

SCIENCES ET TECHNIQUES DU LANGUEDOC

THÈSE

Présentée par

François ALLAL

Pour obtenir le grade de

DOCTEUR

Section C.N.U. : 67 Biologie des populations et écologie

Formation Doctorale : Évolution, Écologie, Ressources génétiques, Paléontologie

École Doctorale : SIBAGHE

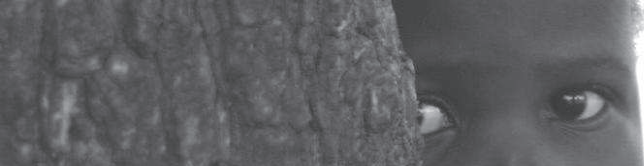


**PATRONS DE VARIABILITÉ CHEZ *VITELLARIA PARADOXA* (KARITÉ) :
ÉTUDE PHYLOGÉOGRAPHIQUE ET ANALYSE COMBINÉE DE LA VARIATION
DES ACIDES GRAS, DES TOCOPHÉROLS ET DE GÈNES CANDIDATS**

Soutenue le 17 Décembre 2010

Devant la commission d'examen formée de :

Mr Joël CUGUEN	Professeur Université de Lille 1	Rapporteur
Mr Thierry ROBERT	Enseignant Chercheur Université Paris-Sud	Rapporteur
Mr Doyle MCKEY	Professeur Université de Montpellier II	Examineur
Mr Philippe GRANDCOLAS	Directeur CNRS MNHN – CNRS	Examineur
Mr François LEFÈVRE	Directeur de Recherche INRA	Co-directeur de thèse
Mr Jean-Marc BOUVET	Directeur de recherche CIRAD	Directeur de thèse



Remerciements

Voici venu la fin d'une belle, longue et périlleuse aventure ! Voici venu le temps de rassembler les premiers éléments de cette expérience, tant sur le plan professionnel que sur le plan humain.

Je manque de recul pour apprécier pleinement tous les bénéfices accumulés au cours de ces trois années de vie. Une chose est sûre, cette expérience unique restera déterminante dans mon développement personnel, au-delà même de sa composante purement scientifique.

Dans cette course de fond et d'obstacles, les questionnements posés et la démarche scientifique m'ont permis d'affirmer ma rigueur et ma capacité de réflexion, et ont constamment suscité et stimulé ma curiosité. Quant aux difficultés rencontrées, si elles ont parfois mis mes nerfs à rude épreuve, elles ont finalement contribué à renforcer mon autonomie et ma maturité.

Cette thèse cependant n'aurait été possible sans le concours de nombreuses personnes...

Mes premières pensées vont tout d'abord à mon directeur de thèse, Jean-Marc Bouvet. Il m'a accueilli dans son unité depuis plusieurs années, lors de mes stages de Masters tout d'abord, puis il m'a accompagné étroitement au cours du déroulement de cette thèse. Merci, Jean-Marc pour cette confiance que tu m'as toujours témoignée. Merci pour cette liberté que tu m'as accordée, qui m'a permis de prendre des initiatives sur le plan scientifique, avec le développement de nouvelles activités au laboratoire. Merci également, et avant tout, pour ta rigueur méthodologique, la grande qualité de ta réflexion intellectuelle, ainsi que le goût et la volonté que tu as de les partager et de les transmettre. Tu as toujours su soutenir et fédérer notre équipe. Même si je me dois de respecter cette sobriété qui est la tienne, je tiens particulièrement à te remercier pour ton grand dévouement. Je profite d'ailleurs de cette occasion pour présenter mes excuses à ta famille, eu égard à toutes les soirées et les weekends que tu m'as consacrés. Merci.

Je remercie également mon co-directeur de thèse François Lefèvre, qui a su également me faire confiance et apporter des perspectives scientifiques déterminantes pour cette thèse et pour la rédaction de l'article « phylogéographie ».

J'adresse mes remerciements les plus chaleureux, et ma grande amitié à Alexandre Vaillant, sans qui, ce travail de thèse n'aurait pu aboutir. Merci d'avoir su partager avec moi les doutes, les joies et les surprises. Tu m'as aidé à affronter les obstacles. C'est ensemble que nous les avons surmontés ... au laboratoire, avec cette espèce qui ne nous aura pas encore livré toutes ses secrets, mais aussi et surtout, au cours de notre mission de récolte au Ghana...

Je remercie tout particulièrement Bénédicte Favreau, qui a su se rendre disponible à chaque instant, au cours de mes manip au laboratoire. Tes conseils avisés sur le plan scientifique ou sur le plan humain m'ont été précieux.

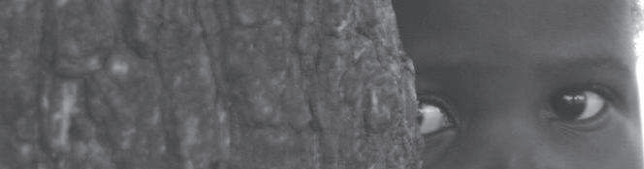
Merci infiniment à Létizia Camus-Kulandaivelu. Merci pour la qualité de ta réflexion scientifique et l'aide que tu m'as apportée sur le plan analytique. Merci pour ta patience et ta pertinence dans les multiples relectures de ce manuscrit. Cet enthousiasme extraordinaire dont tu sais faire preuve a été très réconfortant et stimulant.

J'adresse également mes remerciements les plus sincères et ma plus grande amitié à Fabrice Davrieux. Merci d'avoir été si disponible lors de mon passage à la maison de la Technologie, tu m'as fait partager ton expérience et ta passion pour la chimiométrie dans un climat de bonne humeur qui est ta marque personnelle. Je suis très sensible à cette confiance dont tu m'as honorée.

Merci également à toutes ces personnes avec qui j'ai eu le plus grand plaisir de travailler pour la caractérisation chimique des fruits. Merci à Georges Piombo, à Nathalie Barouh, à Bruno Baréa et à Pierre Villeneuve de m'avoir accueilli et conseillé.

Joël Cuguen et Thierry Robert, ayant accepté d'être les rapporteurs de ce travail de thèse, je vous remercie grandement pour l'intérêt que vous portez à ce manuscrit.

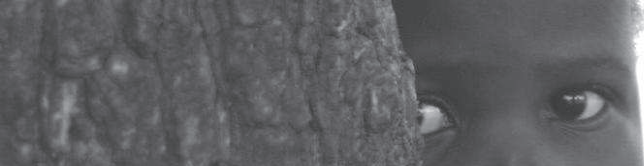
J'adresse maintenant un immense merci, à toute l'équipe de l'UR39 du CIRAD. En particulier, merci au petit Laurent, parti sur le front Pacifique, avec qui j'ai partagé des grands moments de recherche. Merci à toi pour ta collaboration dans l'étude de



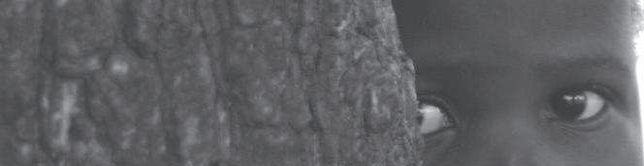
phylogéographie. Merci Zénor Logossa, tu m'as longuement supporté et j'ai eu un grand plaisir à collaborer avec toi, lors des expérimentations et des réflexions analytiques menées tout au long de ce travail de thèse. Merci également à Anna, à Dyana, à Haïchatu, à Marie, à Daniel, à Jean-Marc (Gion), à Alba et à Ronald. Je ne peux clôturer mes remerciements sans exprimer ma gratitude envers Roselyne Lannes, qui, en sa qualité d'assistante a su apporter des solutions à tout problème administratif, mais surtout prêter une oreille attentive à toutes les difficultés humaines.

Cette expérience n'aurait pas non plus été possible sans le soutien sans borne de toute ma famille. À ce titre j'adresse mes remerciements et ma plus grande affection à ma mère pour son dévouement à toutes épreuves et ses conseils rédactionnels avisés, merci maman. Merci encore à mon père, à mon frère jumeau, Pierre, à mes sœurs adorées, Louisa, Sarah et Victoria, et à ma marraine Jacqueline. Merci enfin, et surtout, à ma petite Candice. Tu « as fait au mieux » pour me supporter jusqu'à la fin de cette thèse. Merci pour ta générosité, merci pour ton enthousiasme. Tu as su mieux que quiconque me rassurer, m'encourager et me faire rire dans les moments les plus difficiles.

Je dédicace cette thèse à Simon, à Charles et à Jean



Résumé & Abstract



RÉSUMÉ. Les patrons de variation au sein des espèces d'arbres résultent d'évolutions complexes dont certaines sont particulièrement liées à la condition d'arbre. Si les facteurs expliquant cette variabilité ont été abordés pour les espèces des zones tempérées, ils restent encore peu connus pour des espèces fruitières des zones tropicales sèches offrant des particularités sur le plan adaptatif et relevant d'un processus de domestication ancien. Pour répondre à cet enjeu, l'objectif scientifique de cette thèse est de mieux comprendre l'impact des facteurs évolutifs sur la variabilité au sein d'une espèce d'arbre en zone soudano-sahélienne. Nous nous sommes intéressés à *Vitellaria paradoxa* (karité), une espèce jouant un rôle économique majeur par la production de beurre à partir des graines, dont l'aire de distribution s'étend du Sénégal à l'Ouganda sous la forme de peuplements naturels et agroforestiers.

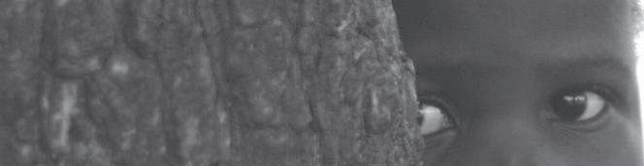
Dans une approche phylogéographique basée sur l'analyse du polymorphisme de marqueurs moléculaires supposés neutres (microsatellites chloroplastiques et nucléaires, et régions intergéniques du chloroplaste) et sur la modélisation des niches écologiques du Karité à différentes périodes du Quaternaire, nous mettons en évidence l'impact des perturbations liées aux dernières glaciations (il y a 20000 ans) sur la diversité génétique de l'espèce. Dans une seconde approche méthodologique, nous étudions l'impact des déterminants environnementaux sur la variabilité des constituants chimiques des graines de Karité, permettant d'énoncer de nouvelles hypothèses. Enfin, nous nous intéressons à l'expression du polymorphisme de gènes candidats, codant les enzymes stéaroyl-ACP-désaturase (SAD) et homogentisate phytyltransférase (HPT), en relation avec la variabilité de la composition chimique en acides gras et en tocophérols des graines. Les résultats obtenus permettent notamment de discuter de l'impact de la sélection naturelle, de la domestication et de la dérive génétique, et d'apporter des hypothèses évolutives et fonctionnelles potentielles expliquant la variabilité observée.

Mots clés. *Vitellaria paradoxa*, Phylogéographie, Facteurs climatiques, Acides gras, Tocophérols, Stéaroyl-ACP-Désaturase, Homogentisate phytyltransférase, Gènes candidats, Tests d'association

ABSTRACT. Patterns of variation within tree species are the result of complex evolutionary changes, some of which are particularly related to the condition of trees. If the factors explaining this variability were discussed for species in temperate zones, they remain little known for fruit tree species of dry tropics, which show special adaptive features and result from an old domestication process. To meet this challenge, the scientific purpose of this thesis is to improve our understanding of evolutionary factors affecting the variability within a tree species in the Sudano-Sahelian zone. We got interested in *Vitellaria paradoxa* (Shea tree), a species which plays a major economic role for the production of butter from seeds and whose distribution range extends from Senegal to Uganda.

In a phylogeographic approach based on the analysis of polymorphism supposedly neutral molecular markers (chloroplast and nuclear microsatellites and chloroplast intergenic regions) and based on ecological niche modelling of Shea tree at different periods of the Quaternary, we shape the role of last glacial maximum (20.000 YBP) on the genetic diversity of the species. In a second methodological approach, we study the impact of environmental determinants on the variability of Shea nuts chemical constituents, stating novel hypotheses. Finally, we investigate the polymorphism of candidate genes, encoding enzymes stearoyl-ACP-desaturase (SAD) and homogentisate phytyltransférase (HPT), in connection with the variability of relative fatty acids compositions and tocopherols contents in seeds. Through the results obtained, we discuss the impact of natural selection, domestication and genetic drift, and provide evolutionary and functional hypotheses that potentially explain the variability observed.

Keywords. *Vitellaria paradoxa*, Phylogeography, Climatic factors, Fatty acids, Tocopherols, Stearoyl-ACP-Desaturase, Homogentisate phytyltransferase, Candidate genes, Association analysis



Sommaire

Introduction Générale _____ **1****1 Particularisme lié à la condition d'arbre ?** _____ **3****2 Étude des déterminants de la variabilité des arbres** _____ **5****3 Objectifs scientifiques de la thèse et hypothèses testées** _____ **7****Chapitre I : Le karité dans l'écosystème de Savane Africaine** _____ **10****1 L'écosystème de Savane Africaine** _____ **11**

1.1 Spécificité de la savane Soudano-Sahélienne _____ 15

1.2 Les parcs agroforestiers et le système de jachère en zone soudano-sahélienne _____ 16

2 Le Karité _____ **20**

2.1 Contexte Biologique _____ 20

2.1.1 Taxonomie et Origine _____ 20

2.1.2 Distribution actuelle des populations _____ 23

2.1.3 Variation environnementale _____ 25

2.1.4 Morphologie de *Vitellaria paradoxa* _____ 26

2.1.5 Biologie de l'espèce _____ 27

2.1.6 Le beurre de Karité _____ 31

2.1.7 Impact des pratiques humaines et domestication du Karité _____ 36

2.2 Contexte socio-économique et historique _____ 38

3 Conclusions _____ **39****Chapitre II : Processus historiques et démographiques ayant participé à la diversité génétique du Karité** _____ **42****1 Past climate changes explain the phylogeography of *Vitellaria paradoxa* over Africa** _____ **43**

1.1 Abstract _____ 44

1.2 Introduction _____ 44

1.3 Material and Methods _____ 46

1.3.1 Plant Material and Taxonomy _____ 46

1.3.2 Molecular markers _____ 47

1.3.3 Statistical analyses _____ 51

1.4 Results _____ 55

1.4.1 Chloroplast sequences _____ 55

1.4.2 Chloroplast SSR Markers _____ 56

1.4.3 Nuclear SSR Markers _____ 61

1.4.4 Present and previous theoretical distribution of *Vitellaria paradoxa* _____ 66

1.5 Discussion _____ 67

1.5.1 Genetic diversity of *Vitellaria paradoxa* _____ 67

1.5.2 A phylogeographic pattern driven by climate changes _____ 68

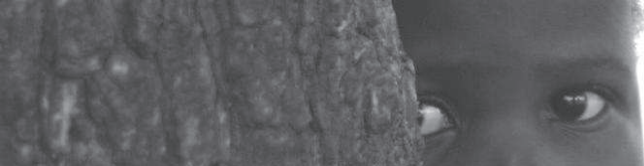
1.5.3 Differentiation between *paradoxa* and *nilotica* _____ 701.5.4 Hypothesis on the origin of *Vitellaria paradoxa* _____ 711.5.5 Outlook for conservation of *Vitellaria paradoxa* in the African savanna biome _____ 71**Box II.1 Detecting Population Expansion in West Africa** _____ **73**

Chapitre III : Variabilité des constituants chimiques au sein des graines du**Karité _____ 75**

1	Introduction	76
1.1	Rôle adaptatif des réserves des graines	76
1.2	Les composés lipidiques des graines	76
1.2.1	Les lipides saponifiables	77
1.2.2	Les lipides insaponifiables	79
1.3	Déterminisme génétique et environnemental de la variabilité des constituants chimiques des graines	82
1.4	Variation en relation avec les gradients environnementaux	83
1.5	Patron de variabilité chez le Karité	84
2	Matériel & Méthodes	86
2.1	Sites échantillonnés, mesures morphologiques et caractéristiques environnementales	86
2.2	Méthodes Analyse chimique	94
2.2.1	Préparation des échantillons	94
2.2.2	Erreur Standard de Laboratoire (ESL)	94
2.2.3	Mesure des teneurs en eau	94
2.2.4	Extraction et mesure des teneurs en Matière Grasse (MG)	94
2.2.5	Préparation des esters méthyliques d'acides gras (EMAG) et analyse des profils d'acides gras (AG)	95
2.2.6	Mesure des teneurs en Tocophérols	96
2.3	Analyses Statistiques	96
2.3.1	Analyses descriptives	97
2.3.2	Analyses de corrélations	97
2.3.3	Analyses des différences entre populations (sites)	97
3	Résultats	99
3.1	Variation morphologique des arbres et leur propagules : l'exemple du Mali	99
3.1.1	Variation générale des mesures morphologiques	99
3.1.2	Variation morphologique des peuplements de Karité du Mali	99
3.2	La teneur en eau	102
3.3	La teneur en matière grasse	103
3.3.1	Variation générale des teneurs en matière grasse	103
3.3.2	Caractérisation des régions et populations de Karité pour la matière grasse	104
3.3.3	Relation entre teneurs en matière grasse et caractères morphologiques	107
3.3.4	Relation avec les facteurs climatiques	107
3.3.5	Variation des teneurs en matière grasse en fonction du type de sol	108
3.4	Les profils en Acides Gras	109
3.4.1	Variation générale des compositions en acides gras	109
3.4.2	Caractérisation des populations de Karité pour les profils d'acides gras	110
3.4.3	Relation avec les caractères morphologiques	116
3.4.4	Relation avec les facteurs climatiques	117
3.5	Les Tocophérols	119
3.5.1	Variation générale des teneurs en tocophérols de la matière grasse	119
3.5.2	Caractérisation des populations de Karité pour les teneurs en tocophérols	120
3.5.3	Relation avec les caractères morphologiques	124
3.5.4	Relation avec les facteurs climatiques	125
4	Discussion	127

4.1	Composition de la matière grasse du Karité par rapport à la littérature	127
4.2	Patron de variation interrégionale des composants lipidiques de la graine de Karité	131
4.3	Patrons de variation intrarégionaux des composants lipidiques de la graine de Karité	133
4.4	Relation avec les variables environnementales	134
4.5	Variations morphologiques et relation avec les constituants chimiques des graines au Mali	135
4.6	Impact de la domestication sur la variabilité chimique	136
5	Conclusion & Perspectives	136
6	Prédiction SPIR de la composition chimique des fruits du Karité	139
Chapitre IV : Étude de la variabilité de gènes candidats		156
1	Introduction	157
1.1	Comment détecte-t-on la signature de la sélection naturelle dans les génomes ?	157
1.2	Bases moléculaires de la variation des Acides Gras (SAD)	162
1.3	Bases moléculaires de la variation des Tocophérols (VTE2-2)	165
1.4	Objectifs et stratégie mise en place dans ce Chapitre	168
2	Matériel & Méthodes	169
2.1	Échantillonnage	169
2.2	Méthodes de laboratoire	170
2.2.1	Comptage chromosomique	170
2.2.2	Extraction de l'ADN génomique	171
2.2.3	Extraction de l'ARN messager	171
2.2.4	RACE (Rapid Amplification of cDNA Ends)	171
2.2.5	Amplification des gènes candidats par PCR	172
2.2.6	Clonage des produits PCR	172
2.2.7	Séquençage	173
2.3	Méthodes d'analyses statistiques	173
2.3.1	Traitement, reconstruction des phases haplotypiques et alignement des données de séquences	173
2.3.2	Analyse du polymorphisme	174
2.3.3	Tests de neutralité	174
2.3.4	Différenciation des populations	176
2.3.5	Constructions des arbres phylogénétiques et des réseaux haplotypiques	176
2.3.6	Mesure du déséquilibre de liaison au sein des gènes	177
2.3.7	Étude d'association entre variants alléliques de VTE2-2 et les teneurs en tocophérol au sein du gradient Nord-Sud au Mali	177
3	Résultats	180
3.1	Étude cytogénétique et comptage chromosomique	180
3.2	Étude du gène SAD chez <i>Vitellaria paradoxa</i>	180
3.2.1	Structure et séquence du gène SAD	180
3.2.2	Étude du polymorphisme des gènes SAD1 et SAD2 chez <i>Vitellaria paradoxa</i>	183
3.2.3	Étude du déséquilibre de liaison (DL) au sein des gènes SAD1 et SAD2	184
3.2.4	Sélection des gènes SAD1 et SAD2 chez <i>Vitellaria paradoxa</i>	187
3.2.5	Différenciation génétique et réseaux des haplotypes des gènes SAD1 et SAD2	190
3.2.6	Analyse phylogénétique des gènes SAD1 et SAD2	195
3.3	Étude du gène VTE2-2 chez <i>Vitellaria paradoxa</i>	198
3.3.1	Structure du gène VTE2-2 chez le Karité	198
3.3.2	Étude du gène VTE2-2 chez <i>Vitellaria paradoxa</i> à l'échelle de l'aire naturelle	198

3.3.3	Étude du gène VTE2-2 chez <i>Vitellaria paradoxa</i> du gradient Nord-Sud au Mali	207
4	Discussion	220
4.1	Le niveau de ploïdie de <i>Vitellaria paradoxa</i>	220
4.2	Une duplication récente du gène <i>SAD</i> chez <i>Vitellaria paradoxa</i>	220
4.3	Une sélection directionnelle des gènes étudiés chez <i>Vitellaria paradoxa</i> à l'échelle de l'aire naturelle	223
4.4	Une adaptation locale de <i>Vitellaria paradoxa</i> au Nord du Mali	226
5	Conclusion et Perspectives	229
	Discussion Générale	231
1	Rappel des principaux résultats et des hypothèses testées	233
1.1	Histoire phylogéographique de <i>Vitellaria paradoxa</i>	233
1.2	Variabilité chimique des graines de Karité	234
1.3	Étude de la variabilité des gènes candidats	235
2	Discussion	237
2.1	La variation au sein de l'espèce et l'impact climatique	237
2.2	Impact de la sélection naturelle et humaine	241
3	Conclusion et Perspectives	243
3.1	Confirmation de résultats connus	243
3.2	Nouveaux résultats et nouvelles hypothèses	243
3.3	Perspectives de recherches	244
3.4	Perspectives de développement	247
	Bibliographie	249
	Annexes	297



LISTE DES TABLEAUX

Tableau I.1.	Variation des dimensions morphologiques des feuilles de <i>Vitellaria paradoxa</i>
Tableau II.1.	Échantillonnage pour l'étude de phylogéographie
Tableau II.2.	Séquence étudiées du génome chloroplastique
Tableau II.3.	Paramètres de diversité au sein de l'échantillon Phylogéographie
Tableau III.1.	Principaux acides gras des graines du Karité.
Tableau III.2.	Description des sites échantillonnés
Tableau III.3.	Températures moyennes et pluviométries mensuelles des sites échantillonnés
Tableau III.4.	Variation des sites selon diverses variables
Tableau III.5.	Statistiques descriptives des mesures morphologiques au Mali
Tableau III.6.	Comparaison des moyennes par site des mesures morphologiques au Mali.
Tableau III.7.	Variation des teneurs en matière grasse par région et par site
Tableau III.8.	Comparaison des moyennes entre régions, sites et transects pour la teneur en matière grasse.
Tableau III.9.	Part relative des différentes sources de variation pour la matière grasse selon différentes échelles
Tableau III.10.	Corrélation entre la teneur en matière grasse et les caractères morphologiques des arbres et des fruits.
Tableau III.11.	Analyse de corrélations entre les teneurs en matière grasse et les variables géographiques, topographiques et climatiques
Tableau III.12.	Comparaison des teneurs moyennes en matière grasse des différents types de sols
Tableau III.13.	Statistiques descriptives pour les profils en acides gras calculé sur l'ensemble de l'échantillon
Tableau III.14.	Pourcentages des différents acides gras et coefficient de variation pour chaque site et chaque région
Tableau III.15.	Corrélation entre les acides gras selon les deux grandes régions
Tableau III.16.	Comparaison des moyennes entre régions, sites et transects pour les deux principaux acides gras
Tableau III.17.	Part relative des différentes sources de variation pour les différents acides gras et selon les différentes échelles
Tableau III.18.	Corrélations entre les acides gras et les critères morphologiques observés

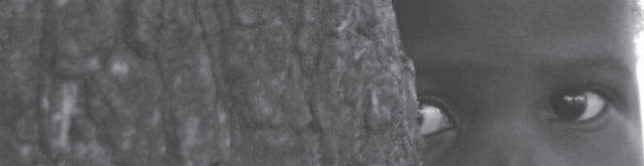
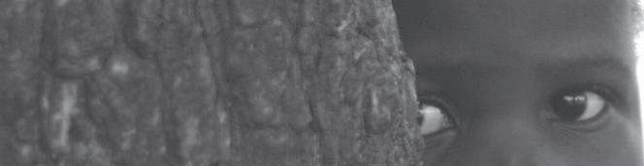
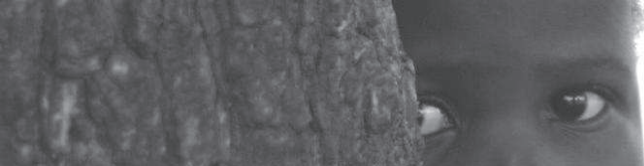


Tableau III.19.	Analyse de corrélation entre les acides gras majoritaires et les variables bioclimatiques
Tableau III.20.	Corrélation entre les deux acides gras majoritaires et les variables climatiques et topographiques au sein des deux régions Ouest et Est
Tableau III.21.	Statistiques descriptives pour les teneurs en tocophérols calculées sur l'ensemble de l'échantillon
Tableau III.22.	Teneurs en tocophérols et coefficient de variation pour chaque site et chaque région
Tableau III.23.	Corrélation entre les formes de tocophérols selon les deux grandes régions
Tableau III.24.	Comparaison des moyennes entre régions et sites pour les deux formes de tocophérols majoritaires
Tableau III.25.	Part relative des différentes sources de variation pour les différentes formes de tocophérols et selon les différentes échelles
Tableau III.26.	Corrélation entre teneurs en tocophérols et variations morphologiques
Tableau III.27.	Corrélations entre tocophérols et variables environnementales
Tableau III.28.	Graphiques de corrélations entre les deux principaux tocophérols et les variables climatiques et topographiques.
Tableau IV.1.	Échantillonnage de l'aire naturelle utilisé pour l'analyse du polymorphisme des gènes candidats.
Tableau IV.2.	Échantillonnage du gradient Nord-Sud au Mali utilisé pour l'analyse du polymorphisme des gènes candidats et les tests d'association
Tableau IV.3.	Nom et caractéristiques des loci microsatellites nucléaires étudiés pour décrire la structure génétique des populations du gradient malien
Tableau IV.4.	Diversité et sélection des gènes <i>SAD1</i> et <i>SAD2</i>
Tableau IV.5.	Niveaux de différenciation génétique entre les populations d'Afrique de l'Ouest, d'Afrique Centrale et d'Afrique de l'Est, pour les différents fragments des gènes <i>SAD1</i> et <i>SAD2</i>
Tableau IV.6.	Diversité et sélection au sein du gène <i>VTE2-2</i> sur l'aire naturelle.
Tableau IV.7.	Niveaux de différenciation génétique du gène <i>VTE2-2</i> entre les populations d'Afrique de l'Ouest, d'Afrique Centrale et d'Afrique de l'Est
Tableau IV.8.	Diversité et sélection au sein du gène <i>VTE2-2</i> au sein du gradient Mali
Tableau IV.9.	Niveaux de différenciation génétique du gène <i>VTE2-2</i> entre les sites du gradient Nord-Sud du Mali
Tableau IV.10.	Paramètres de diversité génétique des sites du gradient Nord-Sud au Mali
Tableau IV.11.	Résultats des tests d'association entre les variants SNP du gène <i>VTE2-2</i> et les teneurs en α -tocophérol, β -tocophérol, γ -tocophérol et δ -tocophérol

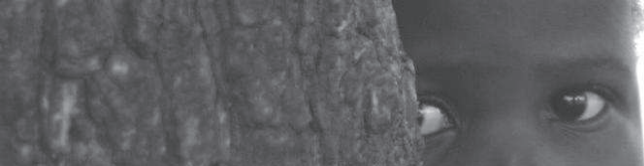


LISTE DES FIGURES

- Figure I.1.** Distribution des écosystèmes de savanes
- Figure I.2.** Facteurs affectant l'abondance de la biomasse ligneuse et herbacée au sein des écosystèmes de savane
- Figure I.3.** Toposéquence en zone de savane au Mali dans la région de Fana
- Figure I.4.** Représentation du système de cycle Culture-Jachère.
- Figure I.5.** Taxonomie de *Vitellaria paradoxa* (Hall *et al.*, 1996)
- Figure I.6.** Distribution géographique de *Vitellaria paradoxa* en relation avec la pluviométrie annuelle
- Figure I.7.** Morphologie de *Vitellaria paradoxa*.
- Figure I.8.** Hiéroglyphiques du tombeau du prince Min-Khaf
- Figure I.9.** Les grandes étapes de la production semi-artisanale de beurre de Karité au Ghana
- Figure II.1.** Relation phylogénétique entre les haplotypes des séquences intergéniques du chloroplaste (Haplotype Network)
- Figure II.2.** Évolution de la richesse haplotypique et de la richesse allélique
- Figure II.3.** Distribution géographique des chlorotypes et représentation de l'arbre phylogénétique basé sur les microsatellites chloroplastiques.
- Figure II.4.** Relation phylogénétique entre les chlorotypes issu des microsatellites chloroplastique (Haplotype Network)
- Figure II.5.** Structure des groupes génétiques reconstruit par le logiciel STRUCTURE (Pritchard *et al.*, 2000) à partir des microsatellites nucléaires
- Figure II.6.** Distribution des groupes de STRUCTURE sur l'aire naturelle
- Figure II.7.** Modélisation des niches écologiques (ENM) à travers trois périodes clés : moderne ; au dernier maximum glaciaire et à la dernière période interglaciaire.
- Figure II.8.** Scenarios testé par DIYABC (Cornuet *et al.*, 2008)
- Figure III.1.** Voie de biosynthèse simplifiée des lipides
- Figure III.2.** Exemple d'un triglycéride : stéarique-linoléique-palmitique
- Figure III.3.** Structure chimique des tocophérols
- Figure III.4.** Voie de biosynthèse des tocophérols
- Figure III.5.** Distribution des sites échantillonnés
- Figure III.6.** Photographie de sites échantillonnés au Ghana
- Figure III.7.** Histogramme de distribution des teneurs en eau

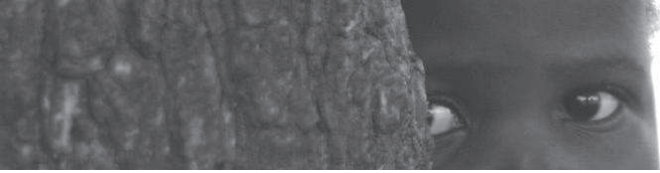


- Figure III.8.** Histogramme de distribution des teneurs en matière grasse
- Figure III.9.** Exemples de chromatogrammes obtenus pour la caractérisation des profils d'acide gras des graines par chromatographie en phase gazeuse.
- Figure III.10.** Histogrammes et corrélations entre les différents acides gras en Afrique de l'Ouest et en Afrique de l'Est
- Figure III.11.** Histogrammes et corrélations entre les différentes formes de tocophérols en Afrique de l'Est et en Afrique de l'Ouest
- Figure III.12.** Comparaison des profils d'acides gras obtenu sur l'aire naturelle dans le projet ProKarité et dans de cette étude
- Figure III.13** Comparaison des teneurs en tocophérols mesurées sur l'aire naturelle dans le projet ProKarité et dans de cette étude
- Figure IV.1.** Impact de la sélection naturelle et des phénomènes démographiques
- Figure IV.2.** Voie de Biosynthèse des acides gras
- Figure IV.3.** Vue tridimensionnelle du dimère protéique de S-ACP-DES chez *Ricinu communis* (Lindqvist *et al.*, 1996)
- Figure IV.4.** Comparaison des morphologies du type sauvage SAD (SSI2) (à gauche) et du type mutant *ssi2* (à droite) chez *Arabidopsis thaliana* (Kachroo *et al.*, 2003a)
- Figure IV.5.** Comparaison des morphologies du type sauvage (au centre) et des mutant *vte2-1* (à gauche) et *vte2-2* (à droite) chez *Arabidopsis thaliana* (Maeda *et al.*, 2006)
- Figure IV.6.** Voie de biosynthèse des tocophérols
- Figure IV.7.** Observation de chromosomes de *Vitellaria paradoxa* en métaphase
- Figure IV.8.** Structure et colinéarité des copies *SAD1* et *SAD2* au sein de la sous-espèce *paradoxa*
- Figure IV.9.** Protéines *SAD1* et *SAD2* chez un individu de la sous espèce *paradoxa*, déduites par traduction *in silico* des séquences d'acide nucléique.
- Figure IV.10.** Déséquilibre de liaison au sein du gène *SAD1*
- Figure IV.11.** Déséquilibre de liaison au sein du gène *SAD2*
- Figure IV.12.** Évolution de la diversité et des indices de sélection du gène *SAD1*
- Figure IV.13.** Évolution de la diversité et des indices de sélection du gène *SAD2*
- Figure IV.14.** Réseau haplotypique de *SAD1_5'*
- Figure IV.15.** Réseau haplotypique de *SAD2_5'*
- Figure IV.16.** Réseau haplotypique de *SAD1_3'*
- Figure IV.17.** Réseau haplotypique de *SAD2_3'*
- Figure IV.18.** Arbre phylogénétique des séquences *SAD1_5'* par Maximum de Vraisemblance.



- Figure IV.19.** Arbre phylogénétique des séquences *SAD2_5'* par Maximum de Vraisemblance
- Figure IV.20.** Arbre phylogénétique des séquences *SAD1_3'* par Maximum de Vraisemblance
- Figure IV.21.** Arbre phylogénétique des séquences *SAD2_3'* par Maximum de Vraisemblance
- Figure IV.22.** Structure et région étudiée au sein du gène *VTE2-2*
- Figure IV.23.** Déséquilibre de liaison au sein du gène *VTE2-2* sur l'aire naturelle
- Figure IV.24.** Évolution de la diversité et des indices de sélection du gène *VTE2-2* à l'échelle de l'aire naturelle
- Figure IV.25.** Réseau haplotypique de *VTE2-2* pour l'échantillon de l'aire naturelle
- Figure IV.26.** Arbre phylogénétique des séquences de *VTE2-2* à l'échelle de l'aire naturelle par Maximum de Vraisemblance
- Figure IV.27.** Déséquilibre de liaison au sein du gène *VTE2-2* sur le gradient Mali
- Figure IV.28.** Évolution de la diversité et des indices de sélection du gène *VTE2-2* au sein du gradient Nord-Sud Malien
- Figure IV.29.** Réseau haplotypique de *VTE2-2* pour gradient Nord-Sud Malien
- Figure IV.30.** Arbre phylogénétique des séquences de *VTE2-2* à l'échelle du gradient Nord-Sud au Mali par Maximum de Vraisemblance
- Figure IV.31.** Structure génétique des populations de *Vitellaria paradoxa* sur le gradient Nord-Sud au Mali
- Figure IV.32.** Distribution des allèles 331:A (en rouge) et 331:G (en vert) au sein de l'échantillon Nord-Sud au Mali
- Figure IV.33.** Variation des valeurs d' α -tocophérol au sein du gradient malien en relation avec les génotypes *VTE2-2_331:A* et *VTE2-2_331:G*
- Figure DG.1.** Prévisions du réchauffement climatique à horizon 2080

Introduction Générale



Afin de répondre aux enjeux récents de conservation de la biodiversité, et plus anciens d'amélioration et de sélection variétale, l'étude de la dynamique de diversité et des déterminants de la variabilité des espèces végétales a connu un essor grandissant depuis la fin du XX^{ème} siècle. Les programmes de recherche les plus avancés se sont intéressés en priorité aux espèces d'intérêt agronomique ayant un rôle majeur dans la sécurité alimentaire (le riz, le blé, le maïs) mais aussi aux espèces menacées patrimoniales (centaurées de la clape par exemple). Les recherches dans ce domaine se sont également attachées aux arbres pour leur rôle dans le maintien des écosystèmes forestiers mais aussi pour l'importance qu'ils occupent pour la production de biomasse pour la construction, pour l'énergie (Zobel & Talbert, 1984) et enfin, pour leur rôle dans la production alimentaire, notamment à travers la production de fruits (Joobeur *et al.*, 1998; King *et al.*, 2000; Ogundiwin *et al.*, 2009).

La compréhension des déterminismes génétiques et environnementaux de la variabilité au sein des espèces d'arbres reste cependant fragmentaire par rapport aux espèces de grande culture. C'est notamment vrai pour les espèces des milieux tropicaux et plus particulièrement pour les arbres fruitiers des savanes africaines qui constituent cependant un fort potentiel sur le plan de la sécurité alimentaire et du développement rural et présentent des particularités originales sur le plan biologique par rapport aux espèces tempérées. C'est dans ce cadre que s'inscrit cette thèse qui porte sur l'analyse des facteurs contribuant à l'expression de la diversité génétique.

Avant de faire un bilan sur ce qui est connu dans le domaine de la génétique des arbres et présenter les objectifs de cette étude, il a semblé avantageux de revenir sur quelques particularités biologiques liées à la condition d'arbres.

1 PARTICULARISME LIÉ À LA CONDITION D'ARBRE ?

Si les arbres ne forment pas en eux-mêmes un groupe naturel sur le plan de la phylogénie, ils partagent des attributs biologiques singuliers dans le règne végétal. Ces plantes vasculaires (trachéophytes) se caractérisent en particulier par une grande taille, une longévité exceptionnelle et un mode de reproduction qui affectent sensiblement leurs patrons d'évolution (Bradshaw, 1972; Petit & Hampe, 2006).

Cette grande taille leur permet de créer un environnement physique et chimique qui influe sur leur propre performance et celle des organismes en interaction (Herwitz *et al.*, 2000; Boege & Marquis, 2005). Cette stature contribue également à limiter les effets des perturbations telles que le pâturage, le piétinement et les feux (Ordóñez *et al.*, 2005), mais rend ces arbres plus sensibles à d'autres perturbations telles que le vent (Loehle, 1988; Gutschick & Bassirirad, 2003; Rowe & Speck, 2005). Le maintien de cette taille nécessite le développement d'une structure tissulaire suffisamment résistante pour supporter la masse aérienne et d'un réseau racinaire performant afin d'assurer la stabilité et un approvisionnement en grandes quantités des ressources. Cette taille entraîne donc des coûts élevés de maintenance, des taux de croissance réduits, et limite de leur existence dans les zones qui n'offrent pas suffisamment de ressources à long terme en énergie, eau et nutriments (Wardle *et al.*, 2004; Ward *et al.*, 2005). Cette grande taille a également pour conséquence l'augmentation des flux de gènes dans le cas des graines et pollens transportés par le vent (Okubo & Levin, 1989; Portnoy & Willson, 1993; Nathan *et al.*, 2002).

La grande longévité permet notamment à ces organismes de supporter des échecs périodiques de reproduction sans conséquences négatives directe sur le plan démographique (Calvo & Horvitz, 1990; Ashman *et al.*, 2004). Cette caractéristique est donc responsable du chevauchement des générations actives sur le plan reproductif, induisant le maintien d'une diversité plus importante que chez les annuelles. Cependant, une longue durée de vie signifie aussi que les individus ont à faire face à des conditions environnementales variables, y compris des événements catastrophiques (Gutschick & Bassirirad, 2003). Par conséquent, cette caractéristique induit chez ces organismes la nécessité d'une grande plasticité phénotypique pour survivre à des cycles annuels successifs et à des changements à plus long terme (Hampe & Bairlein, 2000; Jónsson, 2004).

Le mode de reproduction des arbres est essentiellement allogame. Cette caractéristique de pollinisation croisée est due notamment aux mécanismes d'auto-incompatibilité (Ehrlén & Lehtilä, 2002), à la dépression de consanguinité, et à la dioécie potentielle des organes reproducteurs (Böhle *et al.*, 1996). En particulier, une forte dépression de consanguinité est fréquente chez les arbres (Husband & Schemske, 1996; Sorensen, 1999).

De part ces attributs biologiques particuliers, les arbres sont uniques en ce qu'ils maintiennent des niveaux élevés de diversité, tout en accumulant lentement de nouvelles mutations, et bien qu'étant capables d'adaptations locales rapides (Petit & Hampe, 2006).

Tenant compte de ces caractéristiques inhérentes à la condition des arbres, nous proposons dans la partie suivante d'aborder les grands axes de recherche menés jusqu'à présent afin de mieux appréhender les déterminants de la variabilité au sein de ces espèces.

2 ÉTUDE DES DÉTERMINANTS DE LA VARIABILITÉ DES ARBRES

Les déterminants de la variabilité au sein des espèces forestières ont été abordés selon trois grands axes d'étude :

➤ L'impact des changements climatiques passés

L'étude de l'impact des changements climatiques passés sur l'évolution des aires naturelles constitue, aujourd'hui encore, une thématique très récurrente, en raison notamment du contexte actuel de changement climatique global (IPCC, 2007). Ces études ont abordé les facteurs évolutifs tels que la dérive et la migration avec un focus particulier sur la dernière grande période glaciaire (Petit *et al.*, 1997; Petit *et al.*, 2003). Elles ont surtout concerné les espèces des zones tempérées (Petit *et al.*, 2005; Heuertz *et al.*, 2006; Petit *et al.*, 2008; Heuertz *et al.*, 2010) et ont permis de mettre en évidence des zones refuges et des zones de confluence des routes migratoires, des zones de plus forte diversité, et des zones d'expansion démographique et spatiale notamment dans le cas des espèces boréales (Tarasov *et al.*, 2000; Petit *et al.*, 2003; Heuertz *et al.*, 2004; Petit *et al.*, 2005; Afzal-Rafii & Dodd, 2007; Aitken *et al.*, 2008; Petit *et al.*, 2008). Ce type d'approche a été développé également pour les espèces tropicales en Amérique Latine afin de mieux cerner l'évolution de la forêt amazonienne et de la forêt atlantique (Bonaccorso *et al.*, 2006; Ramos *et al.*, 2007; Dick & Heuertz, 2008; Ramos *et al.*, 2009). Néanmoins, ces études ont été moins développées pour les espèces asiatiques, et très peu de résultats sont disponibles en ce qui concerne les espèces du continent africain.

➤ L'analyse de la variabilité des caractères phénotypiques

D'autres études se sont focalisées particulièrement sur l'analyse de la variabilité des caractères adaptatifs et des caractères d'intérêt agronomique, principalement pour les espèces impliquées dans les schémas d'amélioration. Dans ce cadre, les études se sont basées sur des dispositifs expérimentaux permettant de séparer les effets génétiques des effets environnementaux. Dans ces sites d'expérimentation, croisant les accessions d'arbres et des conditions écologiques contrôlées (température, précipitation, caractéristiques des sols, etc.), ces études ont permis d'identifier les facteurs biotiques et abiotiques responsables des variations observées. Il est cependant à noter que les projets de recherche engagés chez ces espèces pérennes se sont focalisés particulièrement sur des caractères de production de biomasse et sur la résistance au froid et aux parasites, surtout chez les espèces à vocation bois d'œuvre et bois d'énergie (Hurme *et al.*, 1997; Johnsen & Skrøppa, 2000; Garnier-Géré & Ades, 2001; Gonzalez-Martinez *et al.*, 2002; Howe *et al.*, 2003; Bouvet *et al.*, 2005; Savolainen & Pyhäjärvi, 2007). Bien qu'il existe un lien étroit entre la variation morphologique des propagules et la fitness des arbres (Silvertown, 1989), les études similaires sur les espèces fruitières restent moins développées, que ce soit pour les espèces tempérées ou les espèces tropicales.

➤ L'analyse des facteurs biologiques liés à la condition d'arbre

Enfin, un dernier axe d'étude s'est intéressé à l'analyse de la variabilité de ces espèces en relation avec la spécificité des facteurs biologiques liés à la condition d'arbre. Hamrick *et al.* (1991, 1992, 1993), par des études basées sur des marqueurs allozymes, RFLP, AFLP et RAPD, montrent que les espèces ligneuses avec de larges aires géographiques, des systèmes de reproduction allogames, des graines dispersées par le vent ou ingérées par les animaux, présentent plus de diversité et une structure génétique spatiale plus faible que les espèces présentant d'autres attributs biologiques (Hamrick *et al.*, 1991; Hamrick *et al.*, 1992; Hamrick *et al.*, 1993; Hamrick & Godt, 1996). Des études plus récentes jugent ces observations cohérentes avec d'autres attributs biologiques spécifiques des espèces d'arbre tels que les mécanismes d'auto-incompatibilité pollinique, la forte dépression de consanguinité, la diœcie des fleurs (Petit & Hampe, 2006), et également les grandes tailles efficaces de population de ces espèces (Brown *et al.*, 2004; Willyard *et al.*, 2007).

➤ Bilan

Sur le plan scientifique, ces études montrent cependant un manque de connaissance dans certains domaines. En particulier, peu d'entre elles se sont intéressées à la compréhension de l'impact des phénomènes climatiques passés dans certaines zones tropicales et notamment la zone soudano-sahélienne. On note également un grand manque de connaissance sur les déterminants de la diversité liés à la biologie de la reproduction et plus précisément à la morphologie et la constitution chimique des graines d'espèce fruitières non domestiquées. Enfin, si de nombreuses études se sont intéressées au polymorphisme de gènes associés à des caractères liés à la biomasse et à la résistance des espèces ligneuses, on constate un manque de connaissance sur le polymorphisme des gènes impliqués dans les constituants des graines.

Sur un plan lié au développement, ces études ont permis de mieux comprendre globalement cette diversité au sein des espèces d'arbres des zones tempérées et productrices de bois d'œuvre. Elles demeurent cependant très limitées pour des espèces de zones arides notamment en Afrique, et pour des espèces dont l'intérêt économique réside dans la production fruitière. Par ailleurs ces travaux ont surtout concerné des espèces encore sauvages ou récemment impliquées dans des schémas d'amélioration. Elles se sont peu adressées aux espèces ayant subi un processus de semi-domestication.

Afin de combler quelque peu ce manque de connaissance, nous nous sommes intéressés, au cours de cette thèse, aux déterminants de la variabilité chez une espèce d'arbre de la zone soudano-sahélienne en Afrique. Présentons maintenant les objectifs scientifiques et les hypothèses testées.

3 OBJECTIFS SCIENTIFIQUES DE LA THÈSE ET HYPOTHÈSES TESTÉES

L'objectif scientifique général de cette thèse est de mieux comprendre l'impact des facteurs évolutifs sur la variabilité au sein d'une espèce fruitière de savane en zone soudano-sahélienne. Afin de répondre à cet objectif général, nous nous sommes pour cela intéressés à *Vitellaria paradoxa* (karité), une espèce distribuée du Sénégal à l'Ouganda sous forme de peuplements naturels et agroforestiers et jouant un rôle majeur sur le plan du développement économique durable de cette région. Nous avons testé diverses hypothèses

à travers trois démarches complémentaires que sont : l'étude de l'impact des changements climatiques passés ; l'analyse de la variabilité de certains constituants chimiques des graines ; et l'analyse du polymorphisme de gènes candidats associés aux caractères chimiques étudiés.

Dans un **premier chapitre** nous décrivons l'espèce dans l'écosystème de savane africaine en faisant ressortir les réflexions concernant les attributs biologiques et la semi-domestication de cette espèce.

Dans le **second chapitre** de cette thèse, nous proposons de tester une première hypothèse à travers une étude phylogéographique.

Hypothèse 1 : *Les changements climatiques, liés aux dernières grandes glaciations, ont eu un impact marqué sur la distribution de la diversité génétique.*

Pour cela, nous avons utilisé le polymorphisme de divers marqueurs moléculaires, supposés neutres (microsatellites chloroplastiques et nucléaires, et régions intergéniques du chloroplaste), observés au sein de 71 peuplements de Karité répartis sur la totalité de son aire naturelle. Les patrons de diversité observés ont ensuite été mis en relation avec les variations géographiques et climatiques présentes et passées (au dernier maximum glaciaire, il y a 20000 ans ; et à la dernière période interglaciaire, il y a 120000 ans).

Dans le **troisième chapitre** de ce document de thèse, nous nous intéressons à l'étude des déterminants environnementaux de la variabilité des constituants chimiques des graines de Karité afin de tester l'hypothèse suivante :

Hypothèse 2 : *Les variations de constituants chimiques des graines de Karité (acides gras et tocophérols) suivent un cline environnemental (liés principalement à la température et à la pluviométrie).*

À cette fin, utilisant un échantillonnage le long de gradients climatiques marqués, nous proposons une étude de la variation chimique (composition en acides gras et teneurs en tocophérols) et morphologique (taille et poids) des graines, en relation avec les variations climatiques et édaphiques.

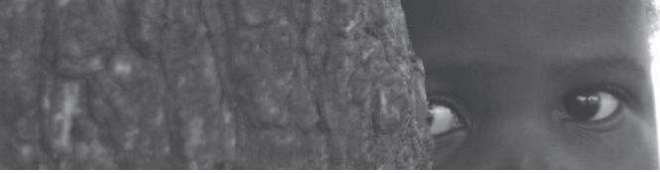
Dans un **quatrième chapitre**, nous testons les hypothèses 3 et 4 à travers l'étude du polymorphisme nucléotidique de gènes candidats impliqués dans les voies de biosynthèses des acides gras et des tocophérols.

Hypothèse 3 : *Le polymorphisme des gènes associés aux constituants chimiques est marqué par l'impact de la sélection naturelle (et/ou artificielle liée à la domestication).*

Hypothèse 4 : *Le polymorphisme des gènes impliqués dans les voies de biosynthèse est associé à la variation des constituants chimiques des graines.*

Pour cela, nous explicitons la structure et le polymorphisme des gènes codant l'enzyme stéaroyl-ACP-désaturase (gène *SAD*) et d'un gène codant pour l'homogentisate phytyltransférase (gène *VTE2-2*), en relation avec la variabilité observée de la composition chimique en acides gras et en tocophérols des graines.

Dans une dernière partie, nous proposons une discussion générale dans laquelle, nous rappellerons les principaux résultats au regard des hypothèses formulées. Nous discuterons les résultats des différentes démarches afin d'énoncer de nouvelles hypothèses et de nouvelles perspectives scientifiques et de développement pour le Karité.



Chapitre I : Le karité dans l'écosystème de Savane Africaine



Afin de comprendre au mieux le contexte de cette thèse, nous dressons dans ce chapitre quelques points bibliographiques importants. Ainsi, dans une première partie, nous décrivons et discutons les caractéristiques majeures de l'écosystème de savane africaine au sein de la région soudano-sahélienne. Dans cette partie, nous parlons en particulier des systèmes de cultures agroforestières, composants fondamentaux de cet écosystème, dans lesquels prédomine le Karité. Enfin, dans une deuxième partie, nous nous intéressons plus spécifiquement au Karité, à travers une discussion des points de connaissance actuels, sur le plan biologique, historique et socio-économique.

1 L'ÉCOSYSTÈME DE SAVANE AFRICAINE

Les savanes sont des écosystèmes comprenant un mélange d'espèces ligneuses (arbres et arbustes) et de plantes herbacées. Elles couvrent environ un cinquième de la surface terrestre mondiale et environ la moitié de la superficie de l'Afrique, de l'Australie et de l'Amérique du Sud (Scholes & Archer, 1997; Sankaran *et al.*, 2004; Sankaran *et al.*, 2005; Beerling & Osborne, 2006) (**Figure I.1**). Concentrant une grande proportion de la population humaine, de la biomasse sauvage (White *et al.*, 2000), de la culture et de l'élevage, les écosystèmes de savane constituent l'un des principaux biomes de notre planète (Sankaran *et al.*, 2005).

Les savanes sont caractérisées par un sous-étage herbacé continu et une couche discontinue d'arbres. Les écosystèmes de savane se caractérisent donc par la coexistence des deux formes végétales que constituent les arbres et les herbacés (Scholes & Archer, 1997). Il est cependant difficile de trouver dans la littérature une définition claire et suffisamment bien délimitée afin de décrire cet écosystème. En outre, Scholes et Archer en donnent une définition pour le moins laconique :

« Savanna occurs where trees and grasses interact to create a biome that is neither grassland nor forest » (Scholes & Archer, 1997).



Chapitre I : Le karité dans l'écosystème de Savane Africaine

La savane apparaît ainsi comme une zone tampon mal déterminée, et peut être considérée comme un intermédiaire entre l'écosystème des prairies (dominance de graminées) et celui de la forêt (dominance des arbres).

Les écosystèmes de savanes sont considérés comme non-équilibrés (Young & Solbrig, 1993), l'abondance relative des plantes herbacées et ligneuses pouvant être très dynamique et évoluer de façon marquée à des échelles de temps décennales (Walker & Noy-Meir, 1982; Archer, 1996). En réalité, les savanes sont constituées d'écosystèmes mixtes très variés, selon la taille des espèces (arbuste vs arbres, herbes hautes vs herbes courtes), selon la dominance ligneuse (à feuilles persistantes ; à feuilles caduques; à feuilles larges ; à feuilles petites ; à racine peu profonde ; ou profondément enracinée), selon le type d'herbacée présente (annuelle ; pérenne ; à cycle photosynthétique en C3 ; à cycle photosynthétique en C4), et selon l'organisation spatiale (distribution aléatoire des arbres ; distribution régulière des arbres...) (House *et al.*, 2003).

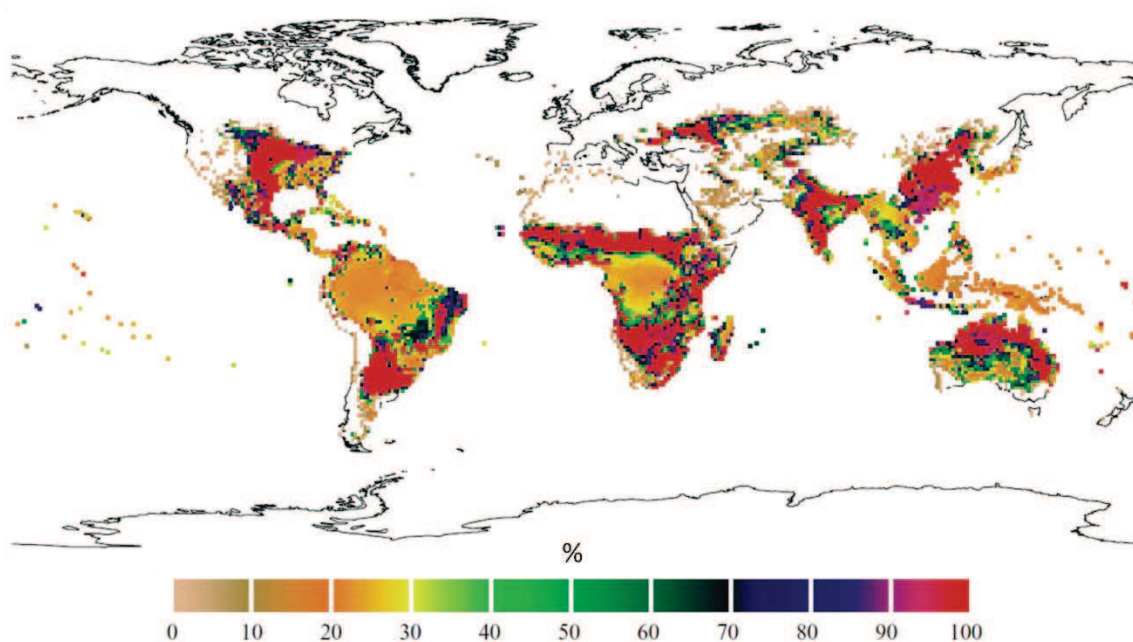


Figure I.1. Distribution globales des proportions relatives des écosystèmes de savanes (Woodward *et al.*, 2004; Beerling & Osborne, 2006)

Ces variations sur le plan fonctionnel et structurel sont le reflet d'un riche éventail d'interactions entre conditions climatiques (en particulier le niveau et la saisonnalité des précipitations), conditions édaphiques (notamment la profondeur et la composition du sol), et pressions anthropiques (en particulier liées à l'élevage, l'occupation du sol et les



Chapitre I : Le karité dans l'écosystème de Savane Africaine

incendies) (**Figure I.2**) (Sarmiento, 1984; Backéus, 1992; Sarmiento, 1992; Scholes & Walker, 1993; Young & Solbrig, 1993).

Ainsi, selon le schéma proposé à la **Figure I.2** et différentes études d'écologie des savanes (Frost, 1986; Boring *et al.*, 1988; Skarpe, 1992; Young & Solbrig, 1993; Van Langevelde *et al.*, 2003; Sankaran *et al.*, 2004; Sankaran *et al.*, 2005), il existe quatre principaux facteurs responsables de la structure et du fonctionnement de ces écosystèmes (House *et al.*, 2003) : la disponibilité en eau des sols (a_w), la disponibilité en nutriments des sols (a_n), l'occurrence des feux de brousse et la prédation des herbivores. La disponibilité hydrique dans les sols dépend de trois facteurs principaux : le régime pluviométrique ; l'évaporation ; et les caractéristiques topographiques et édaphiques (nivellement, composition, et profondeur du sol) (Scholes & Walker, 1993). La quantité et la disponibilité des nutriments du sol dépendent essentiellement des caractéristiques édaphiques, mais résultent en réalité d'une interaction complexe entre le type de sol, le climat et les caractéristiques de la biocénose évoluant dans le milieu (Scholes & Walker, 1993). Si ces deux premières contraintes (a_w et a_n) peuvent être attribuées essentiellement aux caractéristiques climatiques et édaphiques, la survenue de feu de brousse et l'herbivorie sont fortement favorisées par les activités de l'homme et en particulier les pratiques pastorales (Scholes & Walker, 1993). Elles affectent en particulier la densité, la structure et la dynamique des composantes végétales (ligneuses et herbacées) de la savane. Enfin, un dernier facteur important qui n'apparaît pourtant pas clairement dans ce diagramme (**Figure I.2**) est ce que Jouve appelle la pression foncière (Jouve, 1993), que l'on peut également définir par l'occupation du sol et l'aménagement du territoire par l'homme (Frost, 1986; Young & Solbrig, 1993).

La confrontation des conditions hydriques des sols et de la composition du tapis herbacé explique la forte proportion de plantes en C4 dans les savanes (Beerling & Osborne, 2006). Ces espèces sont plus compétitives que les plantes en C3, pour exploiter les ressources limitées des sols de savanes (House *et al.*, 2003; Beerling & Osborne, 2006). En outre, le cycle photosynthétique des plantes C4 est plus efficace (Farquhar *et al.*, 1989; Ehleringer *et al.*, 1991), car notamment, la photosynthèse C4 reste possible malgré une fermeture partielle des stomates, qui intervient dans les milieux à forte radiation solaire afin de limiter l'évapotranspiration des plantes. Les plantes C4 ont également une efficience

accrue dans l'utilisation des ressources hydriques et minérales, car elles ont évolué vers des stratégies « *Water spenders* » d'évitement des contraintes de stress hydrique (Roark & Quisenberry, 1977) en bouclant leur cycle phénologique au cours de la saison des pluies. Cependant, en condition de déficit en azote du sol, les plantes C4 perdent leur avantage sélectif face, en particulier, aux légumineuses à nodulation actinorhizienne leur permettant d'entretenir une symbiose avec des bactéries telluriques fixatrices d'azote atmosphérique du genre *Rhizobium* (Sawada *et al.*, 2003).

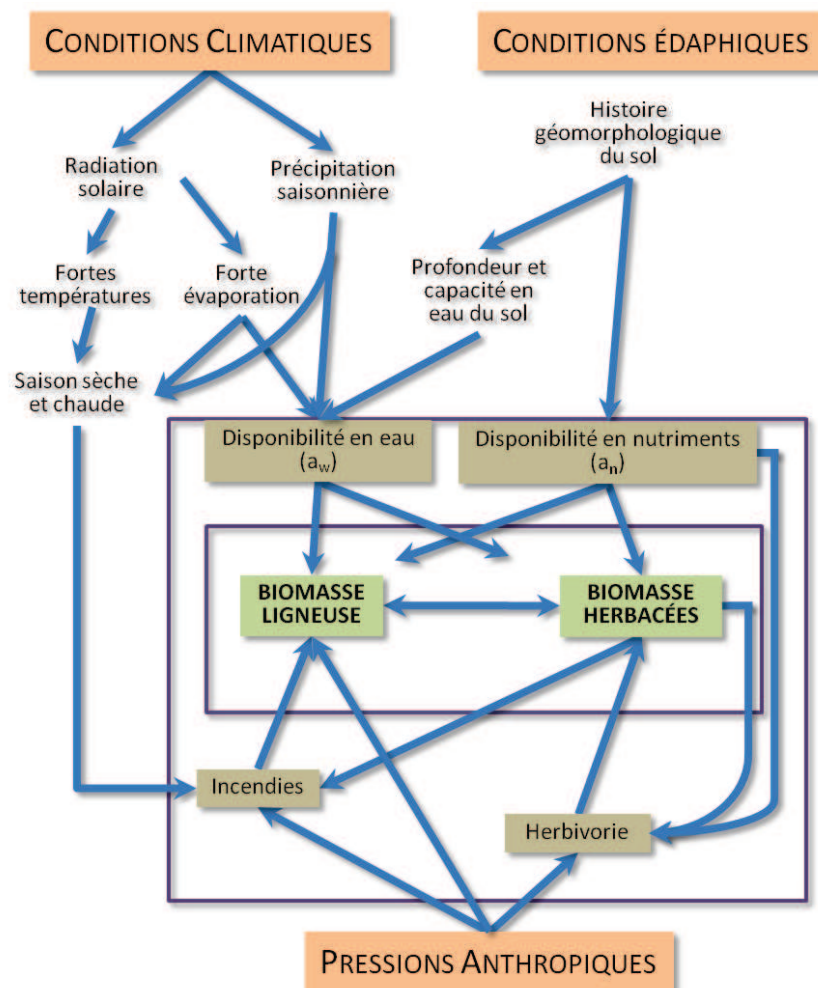


Figure I.2. Facteurs non exhaustifs affectant l'abondance de la biomasse ligneuse et herbacée au sein des écosystèmes de savane (Scholes & Walker, 1993).

La composante ligneuse des écosystèmes de savane possède des exigences sur le plan hydrique très différentes de celles des herbacées (Walker & Noy-Meir, 1982; Scholes & Walker, 1993). En effet, les ligneux disposent de systèmes racinaires profonds et/ou de systèmes superficiels très étendus pour exploiter un plus gros volume de sol, voir les nappes phréatiques. Ces espèces sont particulièrement xéromorphes, ayant développé des organes



adaptés à de grandes périodes sèches. Il ne fait aucun doute que la composante ligneuse des savanes a un rôle primordial dans la création d'îlots de fertilité au sein des écosystèmes de savane, bénéficiant *in fine* au développement des espèces herbacées. En outre, la décomposition de la matière organique, issue de la canopée de l'arbre, enrichit les couches superficielles du sol (Mordelet *et al.*, 1993). De plus, les espèces ligneuses servent d'habitat aux espèces animales et aux insectes, facilitant ainsi la dissémination des graines et des pollens des espèces végétales de l'écosystème (Mordelet *et al.*, 1993).

1.1 Spécificité de la savane Soudano-Sahélienne

La ceinture Soudano-Sahélienne constitue une grande partie de la savane africaine. Elle s'étend d'Est en Ouest sur plus de 6000 km, et se compose d'un ensemble complexe d'écosystèmes de savane bordés par le désert du Sahara au Nord et par les forêts tropicales humides et le golfe de Guinée au Sud. Or cet espace de savane subit en particulier de grands bouleversements depuis quelques décennies (Sankaran *et al.*, 2005; Beerling & Osborne, 2006). De plus, cette zone se caractérise par un accroissement démographique constant des populations humaines, se traduisant par une augmentation des besoins agricoles repoussant sans cesse la limite des terres exploitées (Floret & Pontanier, 2000; Sankaran *et al.*, 2004). En particulier, l'utilisation d'un mode d'agriculture en rotation culture-jachère tend à diminuer, les temps de cycles culture-jachère, indispensables à la régénération du sol, tendent à se raccourcir, et la coupe des espèces ligneuses tend à augmenter pour les besoins en bois combustible/construction et pour les besoins d'extension des espaces de pâturages (Floret & Pontanier, 2000). La dominance au sein de cette région Soudano-Sahélienne de sols peu fertiles, le plus souvent ferrugineux, très sensibles à l'érosion et à la surexploitation agricole, accroît encore la fragilité de ces écosystèmes (Oldeman & Van Engelen, 1993).

Ainsi, il y a une trentaine d'années, la communauté scientifique percevait les écosystèmes de savane comme des biomes capables de faire face à tout bouleversement temporel des conditions climatiques et à toute pratique humaine (Walker & Noy-Meir, 1982). Cependant, face au contexte actuel général que nous venons d'évoquer, Sankaran avec l'appui de la communauté des écologues appelle récemment à mieux comprendre ces écosystèmes complexes sur lesquels pèsent de grandes menaces climatiques, économiques et démographiques :

« Savannahs are anticipated to be among the most sensitive ecosystems to future changes... Understanding of factors that structure savannah communities is urgently required to guide management efforts » (Sankaran et al., 2005)

La valorisation agronomique de cet écosystème de savane soudano-sahélienne s'est organisée traditionnellement à travers la mise en place de parcs agroforestiers dont il est important de décrire les principaux déterminants fonctionnels.

1.2 Les parcs agroforestiers et le système de jachère en zone soudano-sahélienne

Les parcs agroforestiers correspondent donc à des éléments structurels fondamentaux pour les pratiques agropastorales au sein des écosystèmes de savanes de la région soudano-sahélienne. Ces parcs répondent notamment aux caractéristiques climatiques édaphiques et démographiques de cette région, bien qu'ils tendent à être aujourd'hui remplacés par des systèmes de cultures intensives.

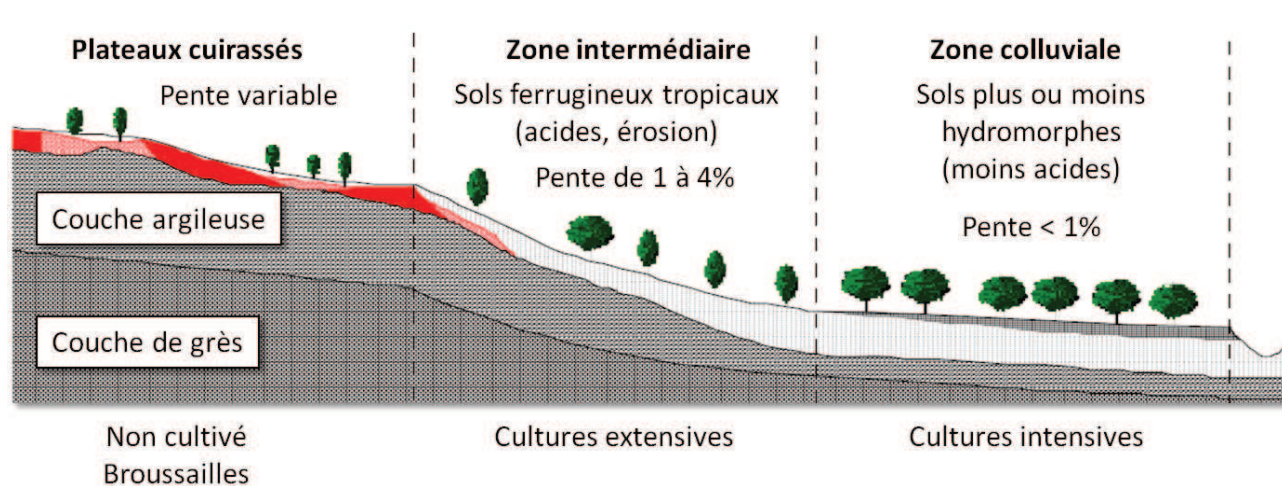


Figure I.3. Exemple d'une toposéquence en zone de savane exploitée au Mali dans la région de Fana (Traoré, 2004)

Ainsi, la zone soudano-sahélienne se caractérise par des températures élevées pendant toute l'année, et un régime pluviométrique à une seule saison pluvieuse, de plus en plus courte quand on se rapproche des tropiques. En raison de la faible couverture végétale dans cette région, les risques d'érosion hydrique et éolienne sont élevés (Traoré, 2004). Parmi les formations typiques des zones arides, on trouve des zones de sols sableux où l'eau



Chapitre I : Le karité dans l'écosystème de Savane Africaine

s'infiltrer généralement bien, des plaines alluviales, des bas fonds et de grands glacis dénudés sur lesquels l'eau ruisselle sans s'infiltrer en totalité (**Figure I.3**).

Dans ce contexte, si l'intensification de l'agriculture en Afrique subsaharienne apparaît nécessaire pour assurer les besoins alimentaires humains, le maintien d'une exploitation rationnelle des terres cultivables est incontournable (Gros, 1979; Johnston & Berg, 1990; Langhelle, 2000). Or comme nous venons de le suggérer, depuis quelques décennies les méthodes de cultures intensives et l'explosion démographique repoussent chaque jour les limites des terres cultivées, le déboisement s'accélère et les rendements agricoles s'amenuisent (Kang *et al.*, 1995; Gillis *et al.*, 1998; Sankaran *et al.*, 2004; Sankaran *et al.*, 2005).

Pourtant, la pratique de l'agroforesterie, basée sur l'alternance de cycles champ-jachère est parfaitement bien adaptée aux écosystèmes tropicaux et très répandue en Afrique sub-saharienne (Floret & Pontanier, 2000). Cette agriculture constituerait en Afrique soudano-sahélienne « *une pratique indispensable à la reproduction des systèmes de culture* » (Jouve, 1993). Cependant, dans le contexte démographique et économique que nous venons de dresser, cette pratique datée du néolithique (Sharma, 1976) est en recul constant en Afrique sub-sahélienne au profit d'agricultures intensives.

Afin de comprendre les déterminants de ces parcs agroforestiers, il est primordial de définir correctement l'intérêt et la place du système de la jachère dans le fonctionnement de ces parcs. La jachère se définit communément comme « *l'état de la terre d'une parcelle entre la récolte d'une culture et le moment de la mise en place de la culture suivante* » (Sebillotte, 1985)...

Période de non-culture, la jachère regroupe en réalité des pratiques très diverses qui n'ont parfois guère de points communs, tant les rôles et fonctions de la jachère sont multiples (Feller *et al.*, 1993; Jouve, 1993). Cependant, quelles que soient les pratiques de jachères utilisées, toutes nous fournissent une clé d'interprétation privilégiée pour comprendre les grands modes d'exploitation agricole du milieu et leur dynamique (Jouve, 1993). Dans les systèmes de production de la zone soudano-sahélienne, la jachère tient deux fonctions primordiales : d'une part assurer l'entretien de la fertilité des sols, d'autre part,

contribuer à l'alimentation des troupeaux. La jachère, si elle est assez longue pour permettre le développement d'un couvert arbustif ou arboré, entraîne un transfert vertical de fertilité, en même temps qu'une extinction des adventices concurrents des cultures.

D'un point de vue général, le système de cycle culture-jachère consiste donc en une phase de culture, suivie d'un abandon cultural (phase de jachère) consécutif à la baisse des rendements (Floret & Pontanier, 2000) (**Figure I.4**). Lors de la phase de culture, dans les systèmes à faibles intrants, les sols subissent une dégradation physico-chimique rapide par une baisse du taux de matière organique, une diminution de la stabilité structurale et une perte des éléments nutritifs (Hoefsloot *et al.*, 1993; Jouve, 1993; Floret & Pontanier, 2000). La durée de la phase d'exploitation d'une parcelle dépend de nombreux facteurs, mais en zone forestière on considère une période de 10 à 15 ans lors de laquelle on observe une production agricole faible mais stable (Floret & Pontanier, 2000). En revanche, la période d'exploitation des zones de savanes herbeuses ne peut excéder 6 ans (Floret & Pontanier, 2000). La phase de jachère qui suit la culture permet le regain de la fertilité du sol grâce à un retour à la savane arbustive ou arborée par régénération naturelle (R.N.) (Floret & Pontanier, 2000). Cependant, on considère que cette remontée biologique est très longue et peut demander jusqu'à 30 à 40 ans (Feller *et al.*, 1993; Floret & Pontanier, 2000). Cette phase d'abandon nécessaire est particulièrement longue, car le milieu en friche devient propice à la survenue de feu de brousse et à l'exploitation de la parcelle sous la forme de pâturage (Floret & Pontanier, 2000).

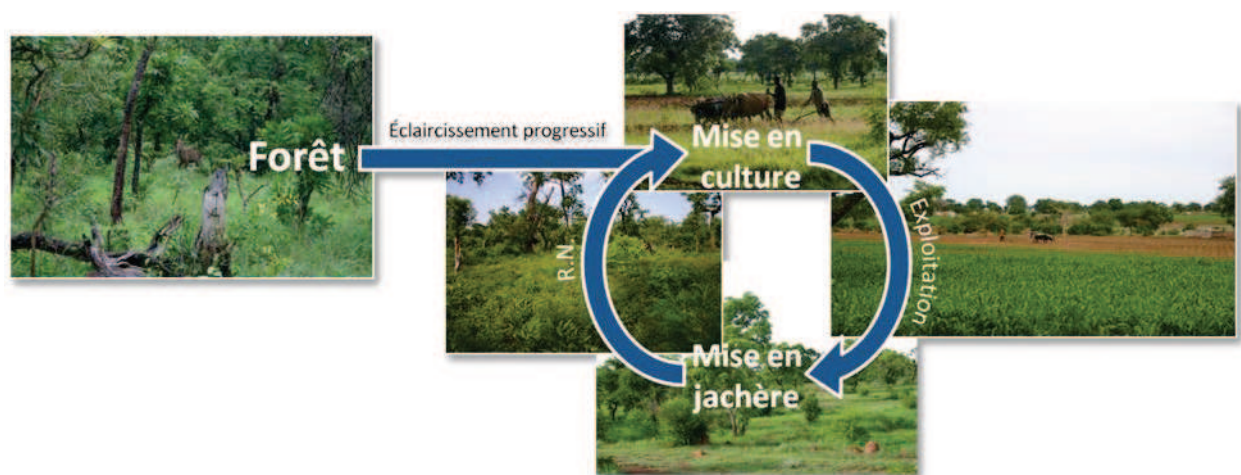


Figure I.4. Représentation du système de cycle Culture-Jachère. R.N. = régénération naturelle.



Chapitre I : Le karité dans l'écosystème de Savane Africaine

Au sein de l'Afrique sub-saharienne, la grande majorité des cultures basées sur ce modèle de rotation, est donc intégrée au sein de parcs agroforestiers. L'agroforesterie repose sur le principe d'une utilisation des terres, dans lequel la composante ligneuse pérenne (arbres, arbustes, arbrisseaux, sous-arbrisseaux, palmiers, bambous) est délibérément cultivée sur des terrains utilisés par ailleurs pour la culture et/ou l'élevage (Raintree, 1987). Les espèces ligneuses conservées dans ces systèmes apportent ainsi un gain biologique indéniable pour les cultures vivrières associées et l'écosystème en général (Rocheleau *et al.*, 1988; Palm, 1995; Schroth, 1998; Jose, 2009): en facilitant la séquestration du carbone ; en permettant le maintien de la biodiversité ; en fertilisant le sol par l'apport de nutriments ; en limitant la dégradation du sol ; et en captant l'eau au niveau de leurs racines profondes. La composante ligneuse apporte également un bénéfice économique important pour les agriculteurs (Jose, 2009). En effet, les arbres conservés sont sélectionnés à des fins diverses, comme par exemple la production de la gomme arabique (*Acacia senegal*) (Anderson *et al.*, 1991), la production du beurre de karité (*Vitellaria paradoxa*) (Teklehaimanot, 2004), ou encore pour la production de bois-énergie (Gruenewald *et al.*, 2007)...

Dans ces systèmes agroforestiers, les interactions économiques et écologiques entre l'élément ligneux et les éléments concernant l'agriculture et/ou l'élevage sont donc très importants (Palm, 1995; Schroth, 1998; Jose, 2009). Et un objectif primordial dans la plupart de ces systèmes agroforestiers consiste à optimiser ces interactions afin de répondre au mieux aux besoins écologiques, économiques et sociologiques (Baumer, 1990).

Dans cette zone soudano-sahélienne de savane, et en particulier dans sa partie occidentale, la composante ligneuse la plus importante au sein de ces systèmes agroforestiers est sans doute le Karité. Nous nous attacherons donc, dans les paragraphes suivants, à dresser le portrait de cette espèce, du point de vue biologique, écologique, historique, économique et social.

2 LE KARITÉ

2.1 Contexte Biologique

2.1.1 Taxonomie et Origine

Le Karité, « *Vitellaria paradoxa* (C.F. GAERTNER) », aussi dénommé « *Butyrospermum parkii* (G. DON) KOTSCHY», appartient à la famille des Sapotaceae de l'ordre des Ébénales (**Figure I.5**). Les Sapotacées constituent l'une des quatre familles existantes au sein de l'ordre des Ébénales, un clade des familles d'angiospermes morphologiquement variable, au sein duquel les relations phylogénétiques sont encore mal comprises (Anderberg *et al.*, 2002).

PHYLUM : Angiospermes
CLASSE : Magnoliopsida
ORDRE : Ébénales
FAMILLE : Sapotaceae
TRIBU : Mimosopeae
GENRE : *Vitellaria*
ESPÈCE : *Vitellaria paradoxa* C.F. GAERTNER
SOUS ESPÈCES : *ssp. paradoxa* et *ssp. nilotica*

Figure I.5. Taxonomie de *Vitellaria paradoxa* (Hall *et al.*, 1996)

En outre, bien que de nombreuses études phylogénétiques se soient intéressées à la place des Sapotaceae au sein des angiospermes (Hutchinson, 1973; Arends, 1976; Takhtajan, 1980; Cronquist, 1981; Dahlgren, 1983; Johnson, 1991; Pennington & Krukoff, 1991), la première reconstruction cladistique de cette famille n'est que très récente dans la littérature (Swenson & Anderberg, 2005).

Les Sapotaceae étaient jusqu'alors divisées en cinq tribus, avec 53 genres et environ 1250 espèces (Pennington & Krukoff, 1991; Govaerts *et al.*, 2001). Les Sapotaceae se composent d'arbres ou d'arbustes avec une distribution dans le monde entier, bien que la plus grande diversité d'espèces se trouve dans les régions tropicales et subtropicales d'Asie et d'Amérique du Sud. Elles se caractérisent par un latex collant et blanc se trouvant souvent dans l'écorce, les branches, les feuilles et les fruits. Elles apparaissent souvent, comme dans les espèces à croissance lente, dans des conditions sèches. Les feuilles sont généralement



Chapitre I : Le karité dans l'écosystème de Savane Africaine

alternes, sauf au sein des genres *Leptostylis*, *Pichonia*, et *Sarcosperma*, présentant des feuilles opposées (Pennington & Krukoff, 1991). La structure des inflorescences fournit de nombreux caractères importants pour la classification au sein des tribus et des genres, et peut être simple ou complexe (Pennington & Krukoff, 1991; Anderberg *et al.*, 2002; Swenson & Anderberg, 2005). Jusqu'à présent, les classifications proposées étaient donc des classifications sur le plan morphologique (Lam, 1939; Aubréville, 1950; Pennington & Krukoff, 1991) dont les résultats étaient souvent contradictoires (Anderberg *et al.*, 2002). Ainsi, Pennington a considéré cinq tribus: les Chrysophylleae, les Isonandreae, les Mimusoepae, les Omphalocarpeae et les Sideroxyleae (Pennington & Krukoff, 1991) et réduit le nombre de genres de 122 à 53. En conséquence, *Chrysophyllum* (81 spp.), *Pouteria* (304 spp.) et *Sideroxylon* (76 spp.) deviennent des groupes difficiles à reconnaître de par une absence apparente de synapomorphies. Swenson et Anderberg suggèrent donc qu'il est urgent de revoir la classification des Sapotacées et proposent trois sous-familles : les Sarcospermatoideae ; les Sapotoideae (comprenant *Vitellaria*), et les Chrysophylloideae (Swenson & Anderberg, 2005).

Les études palynologiques montrent que les premiers pollens de sapotacées apparaissent au Maestrichtien, il y a 65 à 69 millions d'années (Muller, 1981). Conservant les groupes précédents, de grandes controverses demeurent également concernant la divergence de ces tribus (Hall *et al.*, 1996). Ainsi, sur la base du nombre de chromosomes, Arends et Johnson suggèrent que les Mimusoepae (dont *Vitellaria*), les Sideroxyleae et les Isonandreae sont les groupes ayant divergé le plus récemment, tandis que les Chrysophylleae seraient le plus ancien (Arends, 1976; Johnson, 1991). Bokdam considère les Mimusoepae comme étant apparus les plus récemment du fait du caractère charnu de leurs cotylédons (Bokdam, 1977). À l'inverse cependant, Waterman et Mahmoud jugent que la présence de composés phénoliques marque l'état primitif de certains des Mimusoepae (y compris *Vitellaria*) (Waterman & Mahmoud, 1991).

Au sein de l'espèce *Vitellaria paradoxa*, on observe dans la littérature une distinction botanique de deux sous-espèces allopatriques : *Vitellaria paradoxa ssp. paradoxa* en Afrique de l'Ouest et en Afrique centrale ; et *Vitellaria paradoxa ssp. nilotica* en Afrique de l'Est. Cependant, Hall *et al.* (1996) notent que l'origine de ces deux sous espèces continue à soulever de nombreuses incertitudes. Ils soulignent d'une part le manque de données



Chapitre I : Le karité dans l'écosystème de Savane Africaine

cytologiques fiables, surtout pour la sous-espèce *nilotica*, et d'autre part le manque de connaissance des patrons de diversité génétique pour les deux sous-espèces (Hall *et al.*, 1996). Comme seule hypothèse, il est proposé qu'une fracture dans la distribution de l'espèce de part et d'autre de la limite entre les bassins versants du Tchad et du Nil, aurait pu intervenir à un moment donné (Hall *et al.*, 1996). Dans ce contexte très hypothétique, White propose que *nilotica* aurait pu alors être issue de populations plus anciennes isolées dans cette partie Est, et que *paradoxa* constituerait une sous-espèce plus récente issue des populations à l'Ouest de cette fracture (White, 1983). Sans pour autant pouvoir dater cette fracture potentielle, Wickens résume l'information sur l'histoire géomorphologique et climatique du Soudan, indiquant que la distinction entre les bassins versants du Nil et du Tchad n'est apparue qu'il y a environ 10 millions d'années (Wickens, 1976). Cependant, une autre hypothèse serait la présence au Pléistocène, il y a 2,5 millions d'années du méga-lac Tchad, lequel aurait pu constituer un obstacle majeur au sein d'une distribution auparavant continue de l'espèce et avoir été à l'origine de la divergence entre les deux sous-espèces (Wickens, 1976; Hall *et al.*, 1996).

Tableau I.1. Variation des dimensions morphologiques des feuilles de *Vitellaria paradoxa* (Hall *et al.*, 1996).

Sous-espèce	Longueur Feuilles	Largeur Feuilles	Origine	Référence
<i>paradoxa</i>	20	6	Sénégal	(Berhaut, 1967)
	16-22	4-7	Mali	(Vuillet & Africa, 1911)
	9-21	3-5	Côte d'Ivoire	(Chevalier, 1943)
	30	11	Ghana	(Chipp, 1927)
	23	7	Ghana	(Coull, 1928)
	14-15	6-7	Benin	(Chevalier, 1943)
	15-20	6-7	Nigeria	(Lamb, 1941)
	10-25	5-8	Afrique de l'Ouest	(Hutchinson & Dalziel, 1973)
	10-25	4-12	Cameroun	(Vivien, 1990)
	14	5	Repu. Centrafricaine	(Aubréville, 1950)
<i>nilotica</i>	27	9.5	Soudan	(El Ghazali, 1993)
	25	11	Ouganda	(Eggeling, 1951)
	18-22	7	Ouganda	(Langdale-Brown <i>et al.</i> , 1964)
	10-25	4.5-14	Afrique de l'Est	(Hemsley, 1968)

Cependant, il est important de noter que la détermination taxonomique elle-même, entre les deux sous-espèces reste très incertaine dans la littérature (Hall *et al.*, 1996). Ainsi, cette distinction taxonomique se base sur l'observation de légères variations dans les



Chapitre I : Le karité dans l'écosystème de Savane Africaine

dimensions de petits organes végétaux (feuilles, pédicelles, étamines...) compilées à partir de diverses études indépendantes (Hall *et al.*, 1996). Un des éléments majeurs justifiant cette classification botanique est que les feuilles de la sous-espèce *nilotica* seraient plus larges que celles de la sous-espèce *paradoxa*, toutefois cette différenciation n'apparaît pas clairement dans la littérature (**Tableau I.1**) (Hall *et al.*, 1996).

2.1.2 Distribution actuelle des populations

Le karité apparaît en bande, entre le 15ème et le 5ème parallèles, inclinée de l'Ouest à l'Est vers le Sud (Ruyssen, 1957) (**Figure I.6**). Cette bande est globalement située entre les isohyètes de 600 mm et 1400 mm de pluviométrie annuelle (**Figure I.6**), et traverse 18 pays différents du Sénégal à l'Ouganda sur 6000 km d'Est en Ouest et 300 à 1000 km du Nord au Sud (Hall *et al.*, 1996). La sous-espèce *paradoxa* est présente au Sénégal, en Guinée, au Mali, en Côte d'Ivoire, au Ghana, au Burkina Faso, au Togo, au Bénin, au sud du Niger, au Nigéria, au Cameroun, au Tchad et en République Centrafricaine (**Figure I.6**). La sous-espèce *nilotica* se trouve en Éthiopie, au Soudan, en Ouganda et au Congo (Hall *et al.*, 1996). On note que des spécimens d'arbres ont été envoyés par *Kew Botanical Garden* dans l'actuelle République Dominicaine en 1899, et en 1913 ils ont été décrits comme des arbres de 15 m de hauteur, en bonne santé, à la floraison libre, mais à faible fructification (Holland, 1922). L'espèce a également été introduite au Honduras, où elle est connue sous le nom de « tango » (Booth & Wickens, 1988).

Les aires respectives des deux sous-espèces sont dites non chevauchantes, distantes l'une de l'autre de seulement 175 km entre la république Centrafricaine et le Soudan (**Figure I.6**). La limite correspond à l'alternance des bassins hydrographiques du lac Tchad et du Nil (Hall *et al.*, 1996). La sous-espèce *nilotica* se trouvant principalement en Ouganda et au Soudan et ne s'étendant que marginalement au Congo et en Éthiopie, a une aire de distribution beaucoup plus fragmentée et compacte que la sous-espèce *paradoxa* (**Figure I.6**) (Hall *et al.*, 1996).

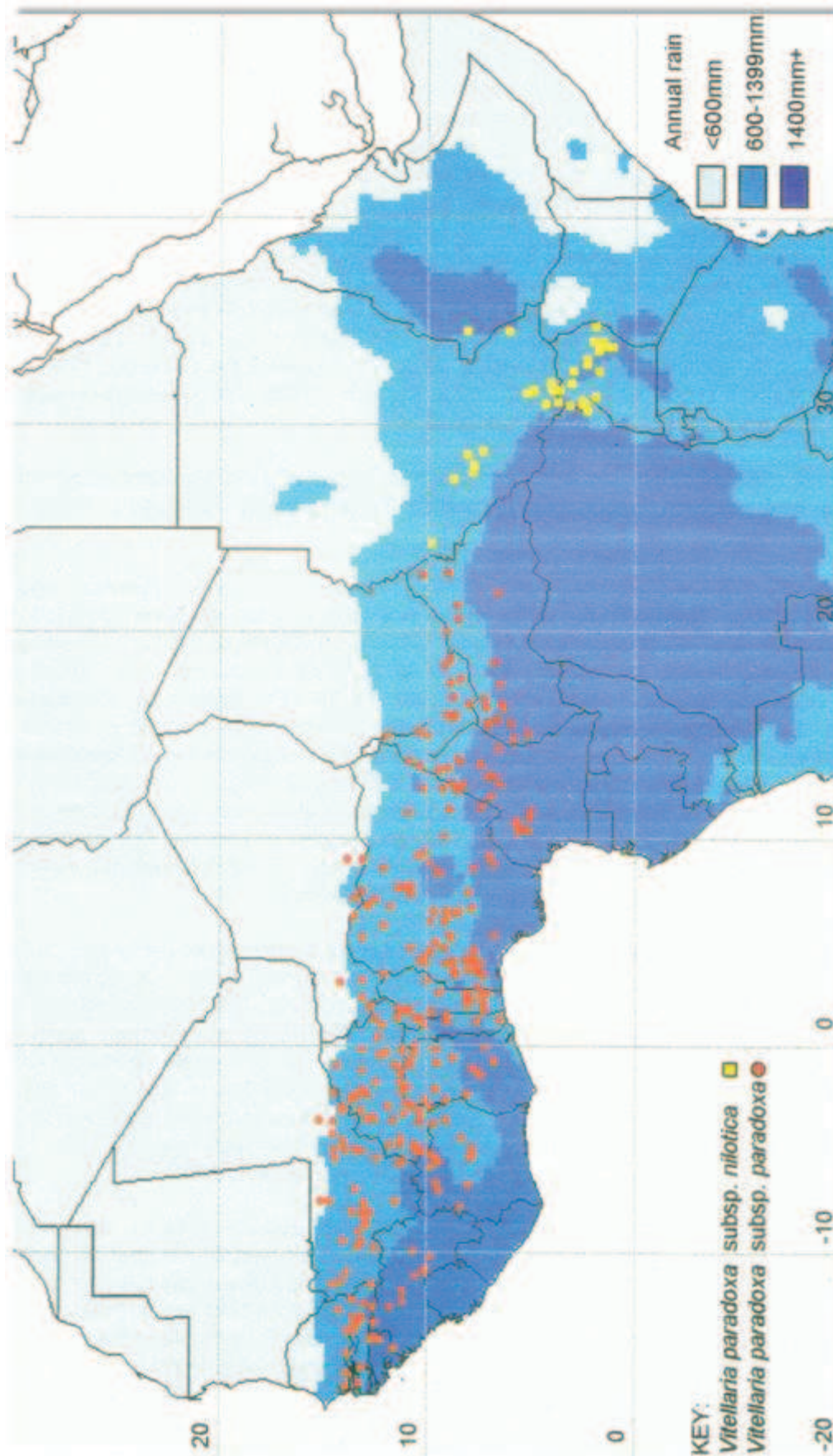


Figure I.6. Distribution géographique de *Vitellaria paradoxa* en relation avec la pluviométrie annuelle (Hall et al., 1996)



2.1.3 Variation environnementale

➤ Altitude

Le Karité se trouve entre 100 et 1600 mètres d'altitude (Hall *et al.*, 1996). Il y a cependant une différence entre les sous-espèces, *nilotica* se situant à des altitudes plus élevées (500-1600 m) que la sous-espèce *paradoxa* (100-1300 m) (Hall *et al.*, 1996).

➤ Climat

Les gammes de variation pluviométrique du karité sont relativement importantes (400 à 1400 mm d'eau par an). On observe des différences entre les deux sous-espèces *paradoxa* (400-1400 mm) et *nilotica* (900-1400 mm) (Hall *et al.*, 1996). Les populations supportent, sur l'aire naturelle, des évaporations potentielles entre 1400 et 2300 mm par an (Hall *et al.*, 1996).

Le nombre de mois secs, définis par des précipitations (en mm) ne dépassant pas 2 fois la température moyenne mensuelle (en °C), varie entre 3 et 8 mois sur l'aire naturelle. Cependant, une fois encore, on observe des différences entre les deux sous-espèces avec seulement 3 à 5 mois secs pour la sous-espèce *nilotica*, contre 4 à 8 mois pour la sous espèce *paradoxa* (Hall *et al.*, 1996).

L'aire naturelle de la sous-espèce *paradoxa* se caractérise en général par une forte température annuelle moyenne (entre 21-29 °C), car l'altitude est la plupart du temps faible (<600 m) (Hall *et al.*, 1996). En revanche, les peuplements de type *paradoxa* situés en plus haute altitude, comme dans le Massif Adamawa au Cameroun, le plateau de Jos au Nigeria ou les hauts plateaux de la Guinée, sont exposés à des conditions plus fraîches avec des températures moyennes annuelles aux environs de 21-22 °C. Concernant la sous-espèce *nilotica*, se trouvant à des altitudes >500 m, la température maximale ne dépasse pas 27-28 °C. Malgré des altitudes comparables à celles des hauts plateaux d'Afrique occidentale pour *paradoxa*, la température annuelle moyenne n'est jamais inférieure à 22 °C dans l'aire orientale, probablement en raison d'un phénomène climatique appelé « Massenerbehung » entraînant un flux d'air chaud sur les reliefs d'Afrique de l'Est (Druitt *et al.*, 1990; Hall *et al.*, 1996).



➤ Géologie et composition des sols

De même que pour la plupart des autres espèces d'arbres tropicaux à grande aire de distribution, on observe une grande variété de types de sol au sein de l'aire naturelle (Hall *et al.*, 1996) (Oldeman & Van Engelen, 1993). On distingue cependant 5 grands types de sols avec du plus pauvre au plus riche : les leptosols, les ferralsols, les acrisols, les lixosols et les vertisols (Oldeman & Van Engelen, 1993). Sur ces sols globalement très ferrugineux, altérés et sensibles à l'érosion, le karité se trouve essentiellement sur des terres composées de grès, de silice, d'argile et de latérites détritiques (Hall *et al.*, 1996).

2.1.4 Morphologie de *Vitellaria paradoxa*

Vitellaria paradoxa est un arbre de taille moyenne (10 à 15 mètres) et montre une grande variabilité sur le plan morphologique, affectant la taille, la forme et la pilosité des feuilles, la couleur, la forme et la taille du fruit et de la graine, l'épaisseur, la consistance et le goût de la pulpe, la période de fructification (Chevalier, 1943; Vivien, 1990; Hall *et al.*, 1996). Le port du karité revêt plusieurs aspects décrits (Chevalier, 1943, 1946; Kelly, 2005). Kelly reconnaît trois types de ports, reconnus également par les paysans, leur prêtant des caractères particuliers de forme de fruits et des feuilles, de précocité ou non (Kelly, 2005). Le port est ainsi en boule (dénommé Borodon au Mali), en balai ou parasol (appelé Boro-boro au Mali), ou intermédiaire (Kolofinian) (Kelly, 2005). La majorité des feuilles de *Vitellaria* apparaît à l'extrémité des branches. On observe parfois jusqu'à 30 feuilles entassées sur les 5 derniers centimètres de la branche (Lely, 1925; Ruysen, 1957; Hall *et al.*, 1996). Cependant, celles-ci peuvent également se développer le long de la branche (Lely, 1925; Hall *et al.*, 1996). La variation des dimensions de feuilles est particulièrement importante sur un même individu (Kelly, 2005; Kelly *et al.*, 2007) (**Figure I.7**). Le **Tableau I.1** résume quelques études portant sur la taille des feuilles sur l'aire naturelle. On distingue cependant deux types de feuilles (Hall *et al.*, 1996; Kelly, 2005; Kelly *et al.*, 2007). Certaines sont petites et claires, très fortement ondulées sur les bords et légèrement acuminées. Les autres sont plus grandes et foncées à ondulations très larges, à sommet arrondi (**Figure I.7**). Les fleurs blanches à couleurs crème du Karité sont hermaphrodites et se développent à l'extrémité des rameaux durant la saison sèche alors même que certaines feuilles persistent (Chevalier, 1943; Hall *et al.*, 1996) (**Figure I.7**). On constate cependant un dimorphisme floral, avec



Chapitre I : Le karité dans l'écosystème de Savane Africaine

l'existence de fleurs longistylées et de fleurs brévistylées (Chevalier, 1943; Hall *et al.*, 1996; Kelly, 2005). Selon les observations de Kelly, les fleurs longistylées seraient fertiles alors que les brévistylées ne le seraient pas (Kelly, 2005).

On distingue sur le plan morphologique trois grandes formes de fruits : oblongue, subglobuleuse et ellipsoïdale (Chevalier, 1943; Eggeling, 1951; Pennington & Krukoff, 1991; Hall *et al.*, 1996) (**Figure I.7**). Les fruits sont rattachés à la branche par un pédicelle de 1.5 à 3 cm de long (Aubréville, 1950; Ruysen, 1957; Hall *et al.*, 1996). L'épicarpe du fruit est vert avant maturation et prend une couleur jaunâtre une fois mûr. Le mésocarpe varie lui aussi de la couleur verte à jaunâtre lors de la maturation (Ruysen, 1957; Hall *et al.*, 1996). La forme de la graine suit généralement celle du fruit (Pennington & Krukoff, 1991). Les graines mesurent, chez les deux sous-espèces, entre 2.5 et 5 cm de long pour 2.2 à 3.5 cm de large (Vuillet & Africa, 1911; Ruysen, 1957; Hemsley, 1968; Vivien, 1990) (**Figure I.7**). La présence fréquente de deux, trois ou rarement quatre graines et plus par fruit peut entraîner de petites variations morphologiques (Chevalier, 1943; Hall *et al.*, 1996; Kelly, 2005). On constate dans les tiges, les feuilles et les fruits la présence de latex (Chevalier, 1943; Ruysen, 1957; Hall *et al.*, 1996; Kelly, 2005).

2.1.5 Biologie de l'espèce

➤ Chromosomes

Miège fait état d'un nombre de chromosomes de $2n = 24$ pour la sous-espèce *paradoxa* pour des spécimens de Niangoloko, au Burkina Faso (Miège, 1954). Ce qui a été confirmé par Pilz pour les échantillons de Kwara, Nigeria (Pilz, 1980).

Les chromosomes varient en taille 0,9 à 1,0 μm à de 2,0 à 2,2 μm (Miège, 1954). Miège note également que la sous-espèce *paradoxa* et une autre Sapotaceae, *Argania spinosa* (L.) Skeels, ont un nombre de chromosomes moins important que les autres membres de la famille, et émet l'hypothèse que la réduction du nombre pourrait être liée à une adaptation aux milieux non-forestiers (Miège, 1954). Des doutes cependant persistent quant au nombre de chromosomes car il semble que l'étude cytologique de *Vitellaria paradoxa* ait été difficile du fait de la petite taille des chromosomes (Miège, 1954; Hall *et al.*,

1996). De plus, comme nous l'avons noté dans les paragraphes précédents, aucune analyse de ce type n'a été effectuée pour la sous-espèce *nilotica* (Hall *et al.*, 1996).



Figure I.7. *Vitellaria paradoxa*. (A) Branche mature ; (A') Graines ; (B) jeune branche ; (B') Tige en floraison ; (a) fleur ; (b) ouverture de la corolle ; (c) coupe verticale montrant l'ovaire supérieur de la fleur ; (d) ovaire pubescent ; (e) coupe verticale d'un ovaire ; (f) coupe transverse d'un ovaire montrant 8 loculi (g) ovule. Gravure (Kotschy, 1865).



Chapitre I : Le karité dans l'écosystème de Savane Africaine

➤ Cycle de reproduction et phénologie

Chez *Vitellaria*, la maturité des fruits est atteinte au début de la saison des pluies, lorsque les conditions climatiques sont les plus favorables à la germination des graines (Hall *et al.*, 1996). Le péricarpe est rapidement éliminé par les insectes ou de petits vertébrés, et, dans des conditions humides, la germination se produit dans les deux semaines consécutives à la chute du fruit (Greenwood, 1929; Ruyssen, 1957). Les graines de *Vitellaria* sont récalcitrantes, leur viabilité diminue rapidement, une semaine après la chute du fruit (Jackson, 1969), et disparaît totalement au bout de 3 à 6 semaines (Sallé *et al.*, 1991).

Dans un classement des membres africains des Sapotaceae, Bokdam réfère *Vitellaria paradoxa* dans une catégorie de plants à cotylédons charnus, se développant à partir de graines rares avec ou sans endosperme (Bokdam, 1977). Les cotylédons charnus, persistants et d'un jaune pâle ou de couleur blanche, montrent des nervations hyphodromes (Bokdam, 1977). L'hypocotyle est absent ou presque, et il n'y a pas d'endosperme. La germination est cryptogée, ce qui représente un moyen important pour survivre aux feux de brousse saisonniers (Bokdam, 1977). Dans les premières années de la croissance des semis, le développement se produit en sous terre. À six mois, on enregistre des hauteurs de tige d'à peine 8 cm avec 3-4 feuilles présentes (Chevalier, 1943; Hall *et al.*, 1996). A 18 mois, ces chiffres ont à peine doublé. On note cependant une croissance racinaire vigoureuse probablement liée à une autre adaptation au feu et à la sécheresse (Hall *et al.*, 1996). On constate une reprise de croissance après le passage des feux de brousse à partir des bourgeons axillaires et à partir de bourgeons à la couronne racinaire (Hall *et al.*, 1996). Cette caractéristique est attribuée à la présence d'une racine pivotante profonde, permettant un stockage d'énergie indispensable aux semis pour produire de nouvelles pousses suite à la destruction de la biomasse aérienne. Ainsi, ce qui peut ressembler à la croissance d'un jeune semi peut en réalité être le produit d'un système racinaire bien développé, vieux de plusieurs années.

La croissance de la tige s'effectue par l'allongement apical du bourgeon terminal jusqu'à l'âge de 6-7 ans, date à laquelle les premières branches commencent à apparaître (Hall *et al.*, 1996). *Vitellaria* est donc un arbre à la croissance particulièrement lente. Le temps nécessaire pour atteindre l'âge de fructification est notamment un frein important à



Chapitre I : Le karité dans l'écosystème de Savane Africaine

la plantation et l'amélioration de cet arbre. Celui-ci n'atteindrait sa maturité reproductive qu'à l'âge de 20 à 25 ans selon les sources (Greenwood, 1929; Chevalier, 1948; Ruyssen, 1957). Cependant, il semble que les arbres ne soient en pleine production que vers l'âge de 50 ans (Greenwood, 1929; Ruyssen, 1957). Eggeling suggère que la sous-espèce *nilotica* atteindrait cette maturité plus rapidement, avec les premières productions de fruits à 12-15 ans et une maturité complète vers 30 ans (Eggeling, 1951). Cependant, Hall *et al.* (1996) suggère de considérer ces données de maturité avec prudence car les auteurs ne semblent par réellement d'accord sur ce point (Hall *et al.*, 1996). C'est également le cas concernant la longévité de *Vitellaria paradoxa*. Cette longévité est estimée entre 200 et 300 ans selon les auteurs (Chevalier, 1948; Ruyssen, 1957).

On constate de grandes variations dans le calendrier des événements phénologiques chez *Vitellaria paradoxa*, tant au sein qu'entre les populations (Hall *et al.*, 1996). La chute des feuilles, la montée de la floraison et la fructification apparaissent comme des événements phénologiques de saison sèche. La chute des feuilles se produit surtout au début de la saison sèche et semble être fortement influencée par le moment des dernières pluies (Chevalier, 1943; Bonkougou, 1987; Hall *et al.*, 1996; Kelly, 2005). Schreckenbergs observe que, dans les rares cas où les arbres sont complètement dépouillés, cet état ne dure jamais plus de deux semaines (Schreckenbergs, 1996). En général, les arbres ne sont jamais complètement défoliés (Coull, 1928; Ruyssen, 1957). Le temps de floraison varie beaucoup également, et peut durer de 30 à 75 jours (Chevalier, 1948). Le début de la floraison intervient généralement à la mi-saison sèche. Ruyssen signale un gradient est-ouest, de la floraison en Afrique occidentale, avec une initiation à la mi-novembre au Togo, par rapport à janvier au Mali (Ruyssen, 1957; Hall *et al.*, 1996). Cependant il a été signalé des populations de *Vitellaria paradoxa paradoxa* à deux floraisons par an, notamment au Nigéria et au Burkina Faso (Hall *et al.*, 1996). La maturation des organes mâles et femelles ne se fait pas simultanément (Kelly, 2005). Les fleurs sont protogynes (maturation des organes femelles avant les mâles), ce qui favorise les fécondations croisées, bien que l'autofécondation ne soit pas impossible (Bonkougou, 1987). La fécondation est donc essentiellement entomophile et assurée principalement par les abeilles (Chevalier, 1943; Bonkougou, 1987; Hall *et al.*, 1996). La coulure est importante car des ombelles pouvant porter 50 fleurs ou davantage ne donnent que rarement plus de 10 fruits. Le développement du fruit dure de quatre à six



mois (Ruyssen, 1957). La fructification est associée au changement de saison sèche à la saison des pluies (Hall *et al.*, 1996). Selon certains auteurs, la floraison et la production de fruits de la sous-espèce *paradoxa* peuvent être cycliques tous les deux à trois ans voire 4 ans (Hall *et al.*, 1996). La production de fruits est ainsi très irrégulière. Un arbre produit en moyenne 15 à 20 kg de fruits frais correspondant à 3 à 4 kg d'amandes sèches. À maturité, les fruits tombent généralement au sol, souvent aidés par des vents violents (Booth & Wickens, 1988) et sont dispersés par l'homme, les éléphants, les oiseaux, les chauves-souris, les rongeurs et les primates (Bonkougou, 1987).

2.1.6 Le beurre de Karité

L'utilisation par l'homme du Karité relève d'une histoire très ancienne. Les premières preuves convaincantes de l'utilisation du Karité, datées de 1000 ans (Ballouche & Neumann, 1995; Neumann *et al.*, 1998) ont été mises en évidence au sein du site archéologique de Saouga au nord du Burkina Faso. La présence de nombreux fragments de coques ainsi que la présence de charbon de bois de karité sur le site, indiquent qu'il est peu probable que ce matériel ait été importé, car seuls sont exportés, le beurre et les graines décortiquées. Le commerce du beurre de Karité est décrit en 1354 par un voyageur arabe, Ibn-Batutta (Lewicki & Johnson, 1974). En 1797, l'explorateur écossais Mungo Park note que les systèmes de culture agroforestiers (décrits précédemment) et l'utilisation du beurre de karité sont très courants en Afrique de l'Ouest sub-saharienne (Park & Rennell, 1799). Des études de linguistique en Afrique de l'Ouest suggèrent une histoire agricole dans la région d'au moins 8000 ans (Ehret, 1984). Cependant, les preuves archéologiques de présence de plantes domestiquées sont rares et les dates sont plus récentes, 5000 ans (Blumler, 1996).

Les écrits de l'Égypte ancienne parlent du commerce régulier avec le pays de « Punt », royaume antique africain englobant l'Ouganda et la corne de l'Afrique (Wicker, 1998). Ce commerce s'effectuait via la grande voie navigable du Nil Blanc, traversant le Sud du Soudan et le Nord de l'Ouganda, où l'huile de karité est actuellement la principale source autochtone d'huile végétale (Masters, 1992; Collinson & Zedie-Bosuener, 1999). Wicker suggère que ce que l'ancienne Égypte appelle l'huile de « hknw » du pays de Punt, serait l'actuel beurre (huile) de Karité provenant de la sous-espèce *nilotica* (Wicker, 1998). Le beurre de Karité aurait-il servi, comme le revendiquent les négociants actuels, à enduire le

corps de la reine Cléopâtre et de la reine de Shaba ? La réponse à cette question n'est pas évidente dans la littérature, mais il semblerait que cette source de d'huile de « hknw » ait eu une importance capitale dans l'ancienne Égypte. Sur les bas-reliefs gravés dans le granite rouge du tombeau du prince Min-Khaf, fils du pharaon Khéops (2600 ans avant JC, Musée du Caire), Smith déchiffre les hiéroglyphes inscrits (**Figure I.8**) :

« Un millier de parfums de festival, un millier d'huile de hknw, un millier d'huile de sft... »



Figure I.8. Extrait des écrits hiéroglyphiques du tombeau de granite rouge du prince Min-Khaf « un millier de parfums de festival, un millier d'huile de hknw, un millier d'huile de sft... » (Smith, 1933)

Cette inscription décrit les offrandes faites au dieu Anubis pour le passage du prince vers l'autre monde (Smith, 1933). L'huile de « hknw », à traduire par « huile sacrée » (Allen, 2000), était placée dans des jarres recouverte d'une fleur de lotus (Pischikova, 2009). Elle aurait ainsi servit d'offrande mais également d'onguent pour l'embaumement des morts. Des doutes subsistent concernant l'huile de « sft », qui pourrait être également de la résine d'un conifère (Allen, 2000).

Actuellement, bien que les techniques traditionnelles de production du beurre de karité varient selon le groupe ethnique et les ressources disponibles, un certain nombre d'étapes distinctes sont communes aux méthodes utilisées par les femmes en Afrique occidentale (**Figure I.9**). Une fois récoltés, les fruits sont dépulpés. Cette pulpe sert alors à l'alimentation humaine ou animale, ou bien épandue, cette pulpe participe à la fertilisation des champs. Les noix sont ensuite généralement bouillies, ce qui permet de désactiver les enzymes de germination d'une part, et de rétracter le noyau pour faciliter de décorticage d'autre part. Ce processus limite également la prolifération de champignons, avant que les noyaux soient séchés au soleil ou par cuisson dans des fours spécialement conçus (Chevalier, 1943; Bonkougou, 1987; Hall *et al.*, 1996) (**Figure I.9**). Une fois que la teneur en humidité a été suffisamment réduite, les coquilles sont éliminées par craquage avec des bâtons ou dans un mortier (**Figure I.9**). Les grains sont alors soigneusement séchés au soleil, au four ou encore fumés jusqu'à ce que la teneur en humidité soit suffisamment faible pour le



Chapitre I : Le karité dans l'écosystème de Savane Africaine

stockage. Les graines sont ensuite stockées pour une période maximale de deux ans dans des silos ou des sacs traditionnels (souvent avec de la cendre) jusqu'à ce que la quantité nécessaire soit atteinte pour la production de l'huile.

Une fois les graines karité séchées et séparées de leurs coques, la procédure d'extraction du beurre peut alors commencer. Les graines sont d'abord écrasées avant d'être torréfiées à sec jusqu'à ce que qu'huile brune soit obtenue (**Figure I.9**). Si aucun broyeur mécanique n'est accessible, les granules grossiers sont pilés dans un mortier en bois pour former une pâte. Après cette étape, il est important de garder la pâte chaude, parce que le beurre de karité se solidifie généralement en dessous de 32 °C (Hall *et al.*, 1996). Après deux heures de battage ou suite au broyage mécanique à chaud, la pâte devient liquide. Elle est alors versée dans une casserole et mise brièvement sur le feu jusqu'à ce que l'huile suinte de la pâte. La pâte obtenue est alors mélangée avec de l'eau pendant environ deux heures jusqu'à ce qu'une émulsion blanche recouvre la surface. Comme cette étape détermine le rendement de l'extraction du beurre, il est important de continuer jusqu'à ce que l'émulsion maximum soit formée.

Plusieurs processus de décantation puis de lavage par chauffage avec de l'eau sont entrepris afin d'obtenir un beurre de la meilleure qualité possible. Le beurre de karité est ensuite laissé à solidifier en une couche blanche homogène à la surface de l'eau refroidie qui plus lourde reste au fond. Le beurre extrait est ensuite récupéré puis conditionné et peut être stocké pendant de longues périodes sans rancir.

Selon la qualité du processus et des graines, de larges gammes de rendements à partir de méthodes traditionnelles d'extraction ont été rapportés, de 10 à 34% du poids du grain (Hall *et al.*, 1996). De nombreuses tentatives ont été faites pour introduire des méthodes semi-industrielles pour l'extraction du beurre de karité, en raison du faible rendement des procédés traditionnels. Cependant, ces processus industriels produisent le plus souvent des rendements et des qualités comparables à ceux des méthodes traditionnelles (Hyman, 1991). Certains acheteurs favorisent donc encore la main traditionnelle (Schreckenber, 1996). La plupart de ces méthodes industrielles utilisent les mêmes méthodes que celles traditionnelles, mais avec des outils mécanisés pour réduire la main-d'œuvre, à savoir l'aide de moulins à moteur (Hyman, 1991).



Figure I.9. Les grandes étapes de la production semi-artisanale de beurre de Karité au Ghana



Chapitre I : Le karité dans l'écosystème de Savane Africaine

Localement, le beurre de karité est destiné à des fins multiples, la plus importante étant son utilisation comme matière grasse alimentaire, car il constitue la source d'huile végétale la moins chère et la plus accessible dans les régions semi-arides d'Afrique subsaharienne (Abbiw, 1995; Lamien *et al.*, 1996; Collinson & Zedie-Bosuener, 1999). D'autres utilisations du beurre de karité sont rapportées : base pour les onguents médicaux et cosmétiques ; pommade ; crème pour les cheveux ; fabrication de savon...

Les changements de réglementation sur les produits de chocolaterie permettent depuis peu l'introduction de 5% de matières grasses végétales autres que de cacao. Il est probable, toutefois, que les Occidentaux mangent souvent du beurre de karité, sans le savoir, car il peut être présent dans de nombreux produits alimentaires, où « huile végétale » est seulement mentionné. Cependant, le marché le plus croissant est celui de l'industrie pharmaceutique et cosmétique. L'industrie cosmétique utilise en particulier le beurre de Karité pour rentrer dans la composition de crème ou de lotion hydratante (Collinson & Zedie-Bosuener, 1999). Le beurre de Karité est également utilisé pour la fabrication de savon.

Les caractéristiques physico-chimiques du beurre de karité ont fait l'objet d'une recherche importante menée notamment par Maranz *et al.* (Maranz & Wiesman, 2003; Maranz *et al.*, 2003; Maranz & Wiesman, 2004; Maranz *et al.*, 2004; Di Vincenzo *et al.*, 2005). Ces études ont permis de montrer que le beurre de karité est essentiellement composé d'acide stéarique, un acide gras saturé à 18 carbones, et d'acide oléique, un acide gras mono-insaturé à 18 carbones également. Elles montrent que les proportions relatives de ces deux acides gras sont très variables sur l'aire naturelle du Karité, induisant des changements importants dans la solidité des beurres extraits. Ces études révèlent également les teneurs très élevées du beurre de karité en matières grasses insaponifiables. Ces composés sont particulièrement recherchés dans les industries cosmétiques et agroalimentaires, car ils sont à l'origine de caractéristiques physico-chimiques importantes. En particulier les tocophérols (ou Vitamine E) sont des composés permettant dans les crèmes d'éviter la dégradation des molécules sous l'effet des processus oxydatifs (Kamal-Eldin & Appelqvist, 1996; Brigelius-Flohé, 2009) et rentrent souvent dans la composition des préparations « anti-âge ». En outre, la variabilité chimique des graines de Karité et donc du beurre extrait est discutée plus en profondeur au cours du Chapitre III de cette thèse.



2.1.7 Impact des pratiques humaines et domestication du Karité

Dans la plupart des essais sur l'écologie de *Vitellaria paradoxa*, cette espèce a le plus souvent été décrite comme «sauvage» par opposition à « domestiquée » (Chevalier, 1943; Baker, 1962; Bonkougou, 1987; Abbiw, 1995; Hall *et al.*, 1996). Cependant, Lovett et Haq (2000a) suggèrent que les karités présents sur les terres agricoles d'Afrique de l'Ouest sont semi-domestiqués, ayant fait l'objet de sélection anthropique à long terme au cours des cycles traditionnels de rotation jachère-culture (Lovett & Haq, 2000a). Un certain nombre d'autres espèces d'arbres africains est également dans un état de semi-domestication. C'est le cas en particulier du baobab (*Adansonia digitata* L.), du balzan (*Acacia albida* (Del.) A. Chev.), du palmier à huile d'Afrique de l'Ouest (*Elaeis guineensis* Jacq.), de l'arbre du chou (*Moringa stenopetala* Bak.F. Cuf.), ou encore du Néré (*Parkia biglobosa* (Jacq.) Don) (Harlan, 1971; Lovett & Haq, 2000a). Il est cependant difficile de distinguer divers stades de domestication en particulier pour les essences d'arbre car ce type de processus est long, continu et itératif. Ainsi, basé sur l'étude de peuplement de Karité au Nord du Ghana et sur des entretiens avec les agriculteurs locaux, Lovett et Haq (2000a) montrent que les pratiques traditionnelles d'agroforesterie ont eu un impact significatif, tant dans la composition des espèces ligneuses, que dans la structure de la population d'arbres. Il apparaît ainsi clairement que la majorité des arbres présents sur les terres cultivées sont maintenues en fonction de leur qualité, évaluée selon les critères locaux (Lovett & Haq, 2000a). Ces processus de sélection interviennent à la fin des périodes de jachères, la plupart des individus immatures sont supprimés et seuls certains grands arbres sont maintenus sur les terres cultivées. Lovett et Haq (2000a) montrent ainsi que, dans les parcelles cultivées, la structure de la population *Vitellaria paradoxa* est différente de celle observée en milieu « naturel » (jamais exploité), avec des proportions plus grandes d'arbres productifs de diamètre, à hauteur de poitrine, supérieur à 20 cm.

Les critères de sélection utilisés sont fondés principalement sur la productivité de fruits, mais également sur des principes d'espacement pour limiter les effets de concurrence avec les cultures annuelles, car cette espèce n'est jamais plantée, les nouveaux arbres étant issus uniquement de la régénération naturelle et des périodes de jachères. Ces critères suggèrent le maintien d'un équilibre fragile entre les cultures vivrières et la productivité des arbres. Lovett et Haq (2000a) proposent également qu'une sélection anthropique



Chapitre I : Le karité dans l'écosystème de Savane Africaine

« inconsciente » ait pu opérer. Ainsi, tenant compte du fait que la totalité des graines non germées sont récoltées, les interdictions rituelles de ramassage pour des durées de 2 semaines (Osei-Amaning, 1996) pourraient avoir encouragé la survie des graines germant rapidement (Lovett & Haq, 2000a). Enfin, la germination cryptogée du Karité pourrait être à l'origine d'une sélection. Les feux provoqués par l'homme pour désherber la parcelle pourraient être à l'origine d'une sélection inconsciente des graines les plus résistantes (Jackson, 1974; Lovett & Haq, 2000a).

Cependant, si Lovett et Haq (2000a) suggèrent que l'espèce est en état de semi-domestication, ils constatent que *Vitellaria paradoxa* conserve cependant beaucoup de caractéristiques indésirables : des arbres arrivant tardivement à maturité (Ruyssen, 1957), des graines récalcitrantes (Hall *et al.*, 1996) et des populations produisant des récoltes très variables en termes de quantité et de qualité (Hall *et al.*, 1996). Ces résultats sont également confirmés par Maranz et Wiesman (2003), qui constatent lors d'une étude de diversité chimique des graines de Karité, des indices de sélections anthropiques sur le plateau de Mossi au Burkina Faso.

Bien que Lovett et Haq (2000a) suggèrent une semi-domestication de l'espèce entreprise depuis des milliers d'années (Lovett & Haq, 2000a), peu d'études ont permis de réellement évaluer l'impact des pratiques humaines sur la variabilité du Karité. Cependant, une étude récente de Kelly *et al.* (2007) révèle que les activités humaines au sein des parcs agroforestiers du Sud du Mali, comme le labourage, la réduction de la densité des arbres, la fertilisation des sols, l'aménagement des terres (champ, jachère et forêt), le mode de gestion paysanne des terres (durée de la jachère), ont un impact considérable sur la phénologie de *Vitellaria paradoxa* (Kelly *et al.*, 2007). Kelly *et al.* constatent en particulier une amélioration de la floraison en champ par rapport aux parcelles de jachères et plus encore par rapport aux peuplements forestiers. Ils proposent ainsi que *Vitellaria paradoxa* profite, en parcelle cultivée, d'une réduction importante de la compétition pour la lumière et les nutriments, du fait de la faible densité d'arbres, de l'élimination des espèces concurrentes, et du bénéfice collatéral des soins prodigués aux cultures associées (Kelly, 2005; Kelly *et al.*, 2007).



Chapitre I : Le karité dans l'écosystème de Savane Africaine

Ces résultats sont également valables pour la production de fruits qui serait moins importante en parcelle de jachère qu'en parcelle de culture (Serpantié, 1996; Lovett & Haq, 2000b), et pour la taille des arbres avec, en particulier, des diamètres plus importants en parcelle cultivée (Kelly *et al.*, 2004a). Enfin, dans une étude récente (Kelly *et al.* en préparation) la jachère apparaît meilleure que la forêt pour l'installation et la croissance de la régénération naturelle grâce au management effectué en période de culture et au mode de gestion (Kelly, 2005). L'alternance jachère-culture constituerait donc un mode de culture bénéfique pour la régénération du Karité.

Ces observations seraient moins évidentes au sein des jeunes jachères (de moins de 10 ans), ce qui renforce l'hypothèse que ce sont les pratiques humaines qui sont responsables de ces patrons (Serpantié, 1996; Kelly, 2005). Kelly *et al.* (2007) notent également que les variations observées, pour la production des fruits, ne peuvent pas uniquement être expliquées par les activités humaines. La phénologie de l'espèce reste encore mal comprise car liée à de nombreux facteurs biotiques, abiotiques et humains (Kelly *et al.*, 2004a; Kelly, 2005; Kelly *et al.*, 2007) difficiles à dissocier.

2.2 Contexte socio-économique et historique

Avec 300.000 tonnes de noix consommées sur le continent Africain, et 300.000 tonnes exportées essentiellement vers les USA et l'Europe, le karité est une des principales sources de matière grasse végétale pour la consommation locale et pour l'exportation (Collinson & Zedie-Bosuener, 1999; Lovett & Haq, 2000b). Le beurre de karité est aujourd'hui utilisé dans l'industrie agroalimentaire (notamment en remplacement du beurre de cacao) et dans l'industrie cosmétique (Collinson & Zedie-Bosuener, 1999). L'estimation par Lovett du tonnage à l'exportation du beurre en Afrique de l'Ouest a été communiquée lors du congrès annuel du *West Africa Trade Hub* en 2008 :

• Bénin	20.000 tonnes
• Burkina Faso	70.000 tonnes
• Côte d'Ivoire	30.000 tonnes
• Ghana	50.000 tonnes
• Guinée-Conakry	7.000 tonnes
• Mali	70,000 tonnes
• Nigeria	40,000 tonnes
• Togo	15,000 tonnes



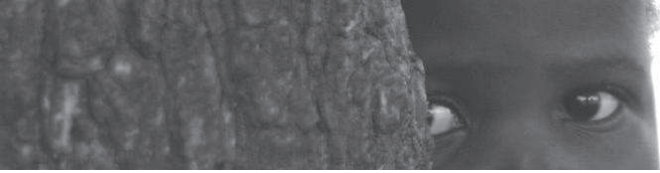
Chapitre I : Le karité dans l'écosystème de Savane Africaine

Cependant, si l'environnement économique autour de cet arbre est important, il est primordial de préciser le contexte social de cette espèce afin de mieux saisir l'impact de cette étude.

Ainsi, comme nous l'avons suggéré dans les paragraphes précédents, la pression d'accroissement démographique et la sédentarisation des populations humaines africaines ont grandement impacté la structure des parcs à karité. L'exploitation de ces arbres est en outre passée d'un système de culture itinérante en cycle culture-jachère, à un système de culture permanente (Hall *et al.*, 1996; Teklehaimanot, 2004; Bayala *et al.*, 2005). Ces changements ont engendré des conséquences néfastes pour les sols et le biotope, en appauvrissant les sols et favorisant l'érosion (Teklehaimanot, 2004). Ainsi, les superficies exploitées sont en telle expansion que des bouleversements sur le plan écologique sont inévitables (Jouve, 1993). Ces évolutions ont un impact sur les ressources naturelles des parcs. Mais cet impact n'est pas nécessairement négatif. Ainsi, pour les parcs à karité, la mise en culture favorise la floraison et la fructification du fait de la réduction de la compétition et de la fertilisation des sols (Kelly *et al.*, 2007). Cependant, les coupes, tailles et mises en jachère (indispensable à la régénération du parc, la jachère facilite l'apparition du feu), les changements climatiques, ont pour conséquences la réduction de la densité des populations, et une mortalité accrue des arbres, surtout dans la partie la plus au Nord de l'Afrique de l'Ouest (Kelly *et al.*, 2004b; Kelly *et al.*, 2007). Il est donc primordial pour la gestion de ces arbres que soient évalués, aussi bien le niveau de la variation génétique que sa distribution. Pour une espèce comme *Vitellaria paradoxa*, il est d'autant plus urgent d'évaluer la diversité génétique, que l'on observe pour cette espèce une réduction de densité, un vieillissement des populations, une baisse de la production, des problèmes de régénération dus aux rigueurs du climat, au feu et à la pression humaine et animale (Kelly *et al.*, 2004b).

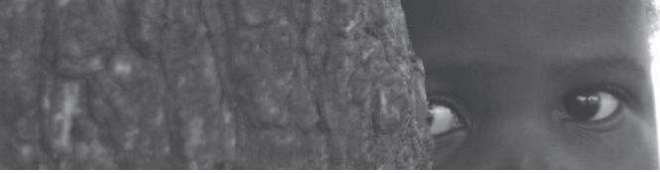
3 CONCLUSIONS

Dans ce chapitre nous avons donc mis en évidence que les savanes africaines sont des écosystèmes fragiles, et particulièrement sensibles au contexte actuel de réchauffement climatique et de pression démographique. Soulignant l'importance du Karité dans la gestion de cet environnement soudano-sahélien et les systèmes agricoles traditionnels

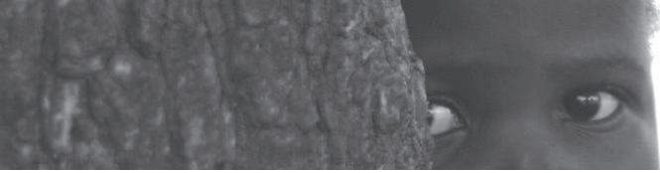


Chapitre I : Le karité dans l'écosystème de Savane Africaine

agroforestiers, nous suggérons, au travers de cette étude bibliographique, que ces systèmes ancestraux de cultures qui constituent une démarche d'avenir et de développement durable, tant que la mise en jachère ne sera pas abandonnée, ont fortement participé à maintenir la diversité des riches écosystèmes de savanes. Pour finir, si nous avons constaté que ces systèmes de cultures et les pratiques humaines associées ont eu un impact important sur la phénologie de l'espèce, nous notons d'une part un manque important de données relatives à la sous-espèce *nilotica*, et d'autre part nous remarquons que très peu d'éléments permettent de comprendre les conséquences de ces pratiques et des facteurs climatiques sur la diversité génétique de *Vitellaria paradoxa*.



Chapitre I : Le karité dans l'écosystème de Savane Africaine



***Chapitre II :
Processus
historiques et
démographiques
ayant participé à la
diversité génétique
du Karité***



Au cours de ce chapitre, nous nous intéressons à l'étude phylogéographique du Karité sur son aire naturelle, par une approche combinant la variation de divers marqueurs : séquences intergéniques du chloroplaste, microsatellites nucléaires et microsatellites chloroplastiques ; et en utilisant des approches récentes de modélisation climatique de l'aire de distribution du Karité, à l'époque moderne, au dernier maximum glaciaire (il y a environ 20000 ans) et durant la dernière période interglaciaire (il y a 120000 ans).

Révélaient un faible patron de différenciation au sein de l'Afrique de l'Ouest, nous abordons, ensuite sous la forme d'un encadré, les résultats d'une étude que nous avons conduite récemment en collaboration avec Zénor Logossa et grâce au concours de son échantillonnage dense dans cette région, afin de mieux comprendre les processus démographiques ayant participé à la diversité génétique du Karité en Afrique occidentale par approche d'ABC (Approximate Bayesian Computation).

1 PAST CLIMATE CHANGES EXPLAIN THE PHYLOGEOGRAPHY OF *VITELLARIA*

PARADOXA OVER AFRICA

Manuscript status: Published in *Heredity* (2011)

Authors names: François Allal¹, Laurent Millet¹, Alexandre Vaillant¹, Létizia Camus-Kulandaivelu¹, Zénor Ablah Logossa^{1,2}, Haby Sanou³, François Lefèvre⁴, Jean-Marc Bouvet¹

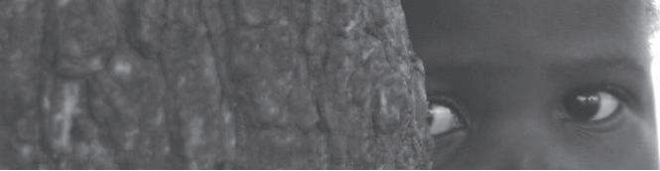
Authors addresses:

¹ Cirad-Bios Department, Research Unit 'Genetic diversity and breeding of forest tree species', international Campus of Baillarguet TA-A 39/C, BP 5035, 34398 Montpellier Cedex 5, France

² University of Lomé, BP: 1515, Lomé, Togo

³ Institut d'Economie Rurale, Forest Resource Programme, Regional Agronomical Research Centre of Sotuba, BP 258, Bamako, Mali

⁴ INRA, URFM, Ecologie des Forêts Méditerranéennes, UR629, Domaine Saint Paul, Site Agroparc, Avignon Cedex 9, France



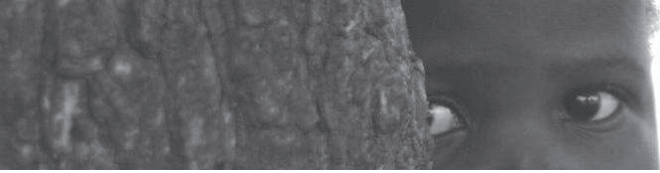
1.1 Abstract

The evolution of the savanna biome has been deeply marked by repeated contraction/expansion phases due to climate perturbations during the Quaternary period. In this study, we investigated the impact of the Last Glacial Maximum (LGM) on the present genetic pattern of *Vitellaria paradoxa* (Shea tree), a major African savanna tree. A range-wide sampling of the species enabled us to sample 374 individuals from 71 populations distributed throughout sub-Saharan Africa. Trees were genotyped using three chloroplast and 12 nuclear microsatellites and were sequenced for two polymorphic chloroplast intergenic spacers. Analyses of genetic diversity and structure were based on frequentist and Bayesian methods. Potential distributions of *Vitellaria paradoxa* today, during the LGM and the last interglacial period were examined using DIVA-GIS ecological niche modelling (ENM). Haplotypic and allelic richness varied significantly across the range according to chloroplast and nuclear microsatellites, which pointed to higher diversity in West Africa. A high but contrasted level of differentiation was revealed among populations, with a clear phylogeographic signal with both nuclear [$F_{ST} = 0.21$; $R_{ST} = 0.28$; $R_{ST} > R_{ST}$ (permuted)] and chloroplast SSRs [$G_{ST} = 0.81$; $N_{ST} = 0.90$; $N_{ST} > N_{ST}$ (permuted)]. We identified a strong geographically-related structure separating western and eastern populations, and a substructure in the eastern part of the area consistent with subspecies distinction. Using ENM, we deduced that perturbations during the LGM fragmented the potential eastern distribution of Shea tree, but not its distribution in West Africa. Our main results suggest that climate variations are the major factor explaining the genetic pattern of *Vitellaria paradoxa*.

Keywords. Intergenic spacers, microsatellites markers, Last Glacial Maximum, phylogeography, sub-Saharan Africa, *Vitellaria paradoxa*.

1.2 Introduction

Today the savanna biome occupies a fifth of the earth's land surface and is characterized by a mixture of grasslands and woodlands (Sankaran *et al.*, 2005). It dates from the late Miocene, about eight million years ago (Cerling *et al.*, 1997; Beerling &



Osborne, 2006). During the Quaternary, the Pleistocene epoch (from 2.5 million to 11,000 years ago) underwent successive cooling and warming periods. During the Last Glacial Maximum (LGM) around 20,000 years ago (Clark *et al.*, 2009), 30% of the Earth's surface was covered by ice, the African continent was drier and the Sahara desert was bigger than today (Lioubimtseva *et al.*, 1998). While many phylogeographic studies have reconstructed the history of holarctic and rain forest species distributions, identifying refugia and routes of postglacial colonization (Bekessy *et al.*, 2002; Petit *et al.*, 2003; Jaramillo-Correa *et al.*, 2008; Petit *et al.*, 2008; Dick, 2010), little attention has been paid to the ecological consequences of these perturbations for savanna ecosystems. Up to now, most studies on this biome have concentrated on animal and plant species in the South American Cerrado, and did not identify a general trend in the impacts of perturbations that occurred during the Quaternary. Confirming paleopalynological studies (Behling, 2002), some savanna species showed evidence of range expansions during the LGM (Bonaccorso *et al.*, 2006), while others showed evidence of fragmentation (Bonaccorso *et al.*, 2006; Ramos *et al.*, 2007; Ramos *et al.*, 2009). Cumulating evidence indicates that plant and animal populations in the African rain forest underwent historical genetic contractions and bottlenecks, surviving in putative refugia during the LGM (Dauby *et al.*, 2010; Lowe *et al.*, 2010), and expanded during postglacial periods to colonize wider areas (Maley, 1989, 1996). In contrast, based on data concerning the Pliocene glacial event 2.5 million years ago (Poumot, 1989; Morley & Richards, 1993), Maley proposes that savanna species underwent similar expansions during the cold and arid periods of the Quaternary (Maley, 1996; Flenley, 1998). To test these assumptions and to help understand the impact of the Last Glacial Maximum on African savanna tree species, we analyzed the range-wide phylogeography of *Vitellaria paradoxa* C.F. Gaertn (commonly known as the Shea tree), an IUCN Red List vulnerable but economically important African savanna tree species exploited in agroforestry systems. Its natural range is a 6000-km wide belt through sub-Saharan Africa covering a contrasted ecological environment. *Vitellaria paradoxa* is currently subdivided in two non-overlapping subspecies, *paradoxa* (from Senegal to Central African Republic), and *nilotica* (in Sudan and Uganda). However, this distinction is based on slight variations in a few morphological traits (the length and width of leaves and stamens) compiled from different studies (Hall *et al.*, 1996) and no data are available concerning their natural or controlled hybridization. Butter



extracted from Shea fruits displays a clear trend, becoming increasingly harder from east to west due to variable ratios of saturated to unsaturated fatty acids (Davrieux *et al.*, 2010).

An initial evaluation of the genetic diversity of *Vitellaria paradoxa* (Fontaine *et al.*, 2004) suggested a genetic differentiation between eastern and western Africa in shaping the genetic structure of the species by emphasizing the role of the Dahomey Gap, a north to south savanna corridor that appeared in the late Holocene (~3,000 years ago) in Togo and Benin, disrupting the West African rain forest cover (Maley 1996). However, this study was performed on a very small sample using RAPD markers, and the results thus need to be considered with caution (Bagley *et al.*, 2001).

To be able to draw general conclusions on the phylogeography of this emblematic African tree, we based our study on a representative sample of the species range using a combination of molecular markers to infer different evolutionary processes. To this end, we used sequence polymorphism of intergenic chloroplast fragments from 58 individuals from 40 populations; and length variation of three chloroplast and 12 nuclear microsatellites genotyped on 364 trees from 71 populations. We used climatic models to assess the distribution of the species in three key periods in the Quaternary era: the modern period, the Last Glacial Maximum, and the last interglacial. Using this experimental design, we tried to answer the following questions: i) what are the impacts of past perturbations on the genetic structure of the species? ii) Can we propose a reliable theory concerning the origin of current populations, with special focus on the differentiation process between the subspecies *paradoxa* and *nilotica*?

1.3 Material and Methods

1.3.1 Plant Material and Taxonomy

Vitellaria paradoxa C.F. Gaertn. (family Sapotaceae), also called Shea tree, is an important savanna tree that is endemic to the African sub-Saharan zone (from Senegal to Uganda). The “butter” extracted from its kernels is the main source of edible vegetable fat in the region. Besides its traditional use, oil is exported throughout Africa and used in the food, confectionary, cosmetic and pharmaceutical industries (Lovett & Haq, 2000). *Vitellaria paradoxa* mostly reproduces sexually and is mainly pollinated by insects, its fruits being



mainly disseminated barochorously and secondarily zoochorously (by birds, monkeys, rodents and humans). Its population density varies greatly depending on land use, locality and ecological conditions. Two subspecies are described in the literature (Hall *et al.*, 1996), ssp. *paradoxa* distributed from Senegal to Central African Republic and ssp. *nilotica*, distributed from the western part of Sudan to eastern Uganda. Their distribution ranges do not overlap and there is a gap of 175 km between the two subspecies due to the different watersheds in the area: the Lake Chad basin is home to *paradoxa* and the Nile drainage basins are home to *nilotica* (Hall *et al.*, 1996).

In the present study, a total of 364 trees were sampled at 71 sites in 14 countries covering the whole distribution range of the two subspecies: Senegal, Mali, Ivory Coast, Ghana, Burkina-Faso, Togo, Benin, Niger, Nigeria, Cameroon, Chad and Central African Republic for *paradoxa*; and Sudan and Uganda for *nilotica*. A subset of samples including 58 individuals from 40 populations was used for the sequence-based chloroplast experiments. Spatial coordinates of populations are given in **Table II.1**. Total genomic DNA was extracted from dried leaves (50 mg) following the Dolezel method (Ky *et al.*, 2000).

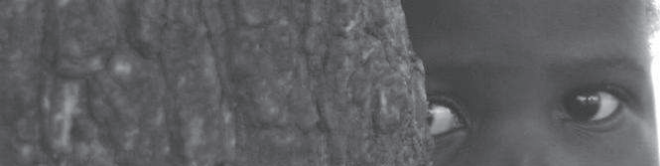
1.3.2 Molecular markers

➤ Genotyping of chloroplast sequences

We successfully amplified and sequenced 11 chloroplast loci on *Vitellaria paradoxa*, including exclusively non-coding intronic or intergenic regions. Universal primer pairs and annealing temperatures are summarized in **Table II.2**. Polymorphism of these loci was assessed in a subset of 10 individuals belonging to 10 different populations including *nilotica* and *paradoxa* subspecies (**Table II.1**). Among these 11 chloroplast loci, which correspond to 7 kb of chloroplast genome, only two presented nucleotide polymorphism: *trnQ-trnS* (Dumolin-Lapegue *et al.*, 1997) and *matK-trnK* spacers (Johnson & Soltis, 1994). These polymorphic loci were sequenced using the above-mentioned subsample (**Table II.1**). Polymerase chain reaction (PCR) was performed in a total of 20 µL, consisting in 20 ng of total DNA, 10 µL of standard QIAGEN PCR buffer, 100 µM of dNTP, 0.2 µM of each primer, and 1 U of Taq DNA polymerase (QIAGEN).

Table II.1. Characterization of sampled populations: Geographic region and country of origin; population name; ID: population code; geographical coordinates; N: number of trees for the different markers; and subspecies.

Region	Country	Site	ID	Longitude	Latitude	Subsp. Species	Nco sequences	NcoSSRs	NnuclSSRs
West Africa	Senegal	Bandafassi	1	W 12.32	N 12.53	ssp. <i>paradoxa</i>	2	5	5
		Samecoula	2	W 12.13	N 12.60	ssp. <i>paradoxa</i>	1	4	4
	Mali	Mahina	3	W 10.85	N 13.78	ssp. <i>paradoxa</i>	1	5	5
		Torodo	4	W 8.84	N 14.50	ssp. <i>paradoxa</i>		4	4
		Badougou	5	W 8.32	N 11.26	ssp. <i>paradoxa</i>	1	4	4
		Kodougouni	6	W 6.66	N 12.97	ssp. <i>paradoxa</i>	1	4	4
		Moeresso *	7	W 5.46	N 12.31	ssp. <i>paradoxa</i>	1	6	6
		Songo	8	W 3.70	N 14.40	ssp. <i>paradoxa</i>	1	2	2
		Tenneli	9	W 3.45	N 14.09	ssp. <i>paradoxa</i>	2	4	4
	Ivory Coast	Napioledougou	10	W 5.57	N 9.28	ssp. <i>paradoxa</i>	2	4	4
		Korhogo	11	W 4.39	N 9.62	ssp. <i>paradoxa</i>	1	2	2
	Burkina-Faso	Peni	12	W 4.45	N 10.97	ssp. <i>paradoxa</i>		6	6
		Karo	13	W 3.30	N 12.36	ssp. <i>paradoxa</i>	1	6	6
		Bonogo *	14	W 1.60	N 12.06	ssp. <i>paradoxa</i>	1	6	6
		Pô	15	W 1.15	N 11.18	ssp. <i>paradoxa</i>	1	6	6
		Konia	16	W 0.99	N 13.10	ssp. <i>paradoxa</i>	1	6	6
		Fada N'Gourma	17	E 0.41	N 12.06	ssp. <i>paradoxa</i>	1	6	6
	Ghana	Larabanga	18	W 1.85	N 9.19	ssp. <i>paradoxa</i>	2	6	6
		Kawampe *	19	W 1.56	N 8.43	ssp. <i>paradoxa</i>	1	6	6
		Kadia	20	W 0.86	N 9.90	ssp. <i>paradoxa</i>	2	6	6
		Yendi	21	W 0.04	N 9.45	ssp. <i>paradoxa</i>	2	6	6
	Togo	Tabango	22	E 0.20	N 10.73	ssp. <i>paradoxa</i>		4	4
		Pavô	23	E 0.69	N 10.23	ssp. <i>paradoxa</i>	1	4	4
		Koudassi	24	E 0.86	N 6.64	ssp. <i>paradoxa</i>		6	6
		Kante *	25	E 1.03	N 9.98	ssp. <i>paradoxa</i>	1	6	6
		Kolina	26	E 1.14	N 9.10	ssp. <i>paradoxa</i>	1	6	6
	Benin-Niger	Diangbassou	27	E 1.19	N 7.81	ssp. <i>paradoxa</i>		6	6
		Cocota *	28	E 1.36	N 10.41	ssp. <i>paradoxa</i>	1	6	6
		Kpale	29	E 1.73	N 9.65	ssp. <i>paradoxa</i>	1	4	4
		Pira	30	E 1.78	N 8.48	ssp. <i>paradoxa</i>	1	6	6
		Kandi	31	E 2.95	N 11.16	ssp. <i>paradoxa</i>	1	6	6
		Kalale	32	E 3.36	N 10.27	ssp. <i>paradoxa</i>	2	6	6
		Boumba	33	E 2.86	N 12.41	ssp. <i>paradoxa</i>		2	2
	Nigeria	Wawa	34	E 4.42	N 9.90	ssp. <i>paradoxa</i>		10	10
		Riaou	35	E 5.25	N 11.10	ssp. <i>paradoxa</i>		10	10
		Paiko	36	E 6.60	N 9.43	ssp. <i>paradoxa</i>		11	11
		Abuia	37	E 7.42	N 9.08	ssp. <i>paradoxa</i>		10	10
		Kuru	38	E 8.81	N 9.68	ssp. <i>paradoxa</i>		6	6
		Shendam	39	E 9.53	N 8.88	ssp. <i>paradoxa</i>		6	6
		Kaltungo	40	E 11.32	N 9.82	ssp. <i>paradoxa</i>		6	6
		Mubi	41	E 13.27	N 10.27	ssp. <i>paradoxa</i>		6	6
	Central Africa	Bangangte	42	E 10.52	N 5.15	ssp. <i>paradoxa</i>	2	6	6
		Foumban	43	E 10.90	N 4.73	ssp. <i>paradoxa</i>	2	6	6
		Ngong	44	E 13.50	N 9.02	ssp. <i>paradoxa</i>		5	5
		Neaoundere	45	E 13.58	N 7.32	ssp. <i>paradoxa</i>		4	4
		Mafa-Kilda *	46	E 13.75	N 9.03	ssp. <i>paradoxa</i>	2	5	5
		Fieuil	47	E 13.95	N 9.75	ssp. <i>paradoxa</i>		5	5
		Mendeo	48	E 14.43	N 10.37	ssp. <i>paradoxa</i>		4	4
		Kelo	49	E 15.78	N 9.20	ssp. <i>paradoxa</i>	1	5	5
	Chad	Mondou	50	E 16.02	N 8.62	ssp. <i>paradoxa</i>	1	5	5
		Sarh *	51	E 18.37	N 9.17	ssp. <i>paradoxa</i>	1	4	4
	Central	North	52	E 20.45	N 8.58	ssp. <i>paradoxa</i>	2	3	3
		Central *	53	E 20.59	N 8.22	ssp. <i>paradoxa</i>	2	6	6
		South	54	E 20.85	N 7.88	ssp. <i>paradoxa</i>	3	6	6
East Africa	Sudan	Kuaiena	55	E 28.25	N 7.52	ssp. <i>nilotica</i>		5	5
		Cuebet	56	E 29.21	N 7.03	ssp. <i>nilotica</i>		5	5
		Wulu	57	E 29.63	N 6.49	ssp. <i>nilotica</i>		3	3
		Kuel Kuac	58	E 29.64	N 6.58	ssp. <i>nilotica</i>		2	2
		Matengai	59	E 29.64	N 6.83	ssp. <i>nilotica</i>		5	5
		Mbara	60	E 30.17	N 5.45	ssp. <i>nilotica</i>		2	2
		Mundri	61	E 30.32	N 5.34	ssp. <i>nilotica</i>		4	4
	Uganda	Arua	62	E 30.90	N 3.03	ssp. <i>nilotica</i>		4	4
		Eria	63	E 31.64	N 3.62	ssp. <i>nilotica</i>		4	4
		Adumara	64	E 33.05	N 2.53	ssp. <i>nilotica</i>	2	4	4
		Otongere *	65	E 33.15	N 2.72	ssp. <i>nilotica</i>	2	4	4
		Okwongo	66	E 33.25	N 2.50	ssp. <i>nilotica</i>	2	5	5
		Lacankweri	67	E 33.31	N 2.80	ssp. <i>nilotica</i>		4	4
		Olwoko	68	E 33.50	N 2.43	ssp. <i>nilotica</i>		5	5
		Otuks ss *	69	E 33.70	N 2.51	ssp. <i>nilotica</i>	2	4	4
		Abarilela	70	E 33.81	N 1.98	ssp. <i>nilotica</i>		5	5
		Akobo	71	E 33.99	N 1.82	ssp. <i>nilotica</i>		4	4



*: populations used for chloroplast polymorphism pre-screening.

Amplifications were performed in a MasterCycler Eppendorf thermocycler (Hamburg, Germany) with an initial 4-min denaturation step at 94 °C, followed by 30 cycles (30 sec of denaturation at 94 °C, 1 min at 61.7 °C for *matK-trnK* and at 68 °C for *trnQ-trnS*, 1 min of extension at 72 °C), then a final extension step at 72 °C for 5 min for *matK-trnK* or for 10 min for *trnQ-trnS*. PCR products were purified and sequenced by GATC Biotech (Konstanz, Germany). Using a similar protocol, we amplified and sequenced *trnQ-trnS* and *matK-trnK* from five individuals of *Baillonella toxisperma* Pierre, a Sapotaceae family relative, used as outgroup species.

➤ Genotyping of chloroplast SSRs

Among the 10 microsatellite primer pairs tested, nine showed PCR amplification on our samples: eight universal microsatellite Ccmp1, Ccmp2, Ccmp3, Ccmp4, Ccmp5, Ccmp6, Ccmp7 and Ccmp9 (Weising & Gardner, 1999); and one tobacco microsatellite, Ntcp9 (Bryan *et al.*, 1999). For determination of polymorphism, these nine microsatellite markers were tested on 10 individuals covering the whole distribution range (**Table II.1**). Three chloroplast microsatellites (cpSSR) were polymorphic (Ccmp3, Ccmp5 and Ntcp9), and were chosen to lead the global analysis of the 364 individuals from 71 populations. The protocol used for amplification and genotyping is described in Müller (Müller *et al.*, 2009) (**see Annexe II.1**).

➤ Genotyping of nuclear SSRs

Among the 16 polymorphic nuclear SSR markers (nucSSR), four were discarded as many genotypes were missing. Finally, 12 polymorphic nuclear SSR markers (Allal *et al.*, 2008) were used (mCIRVp08, mCIRVp14, mCIRVp28, mCIRVp70, mCIRVp113, mCIRVp134, mCIRVp159, mCIRVp167, mCIRVp168, mCIRVp175, mCIRVp181, and mCIRVp191). PCR conditions and genotyping methods were identical to those described by Allal (Allal *et al.*, 2008).

Table II.2. Chloroplast non-coding sequences amplified: locus names; primer sequences; size of the fragments sequenced; hybridization temperature; EMBL accession numbers and references associated with these loci.

Locus	Primers sequences		PCR		EMBL accession N°	
	Forward primer (5'-3')	Reverse primer (5'-3')	Fragment size (bp)	Ta (°C)	<i>paradoxa</i>	<i>nilotica</i> Ref.
<i>matK-trnK</i> (Spacer)	matK8 : CTTCGACTTTCTTGCT	trnK2r : AACTAGTCGGATGGAGTAG	452	61.7	FN908107 FN908109 (Cameroun) FN908108 (Central African Rep.) FN908097 FN908098 (Chad)	FN908110 a
<i>trnQ-trnS</i> (Spacer)	trnQ : GGGACGGAAGGATTGGAACC	trnSr : GGGAGAGATGGCTGAGTGGAC	601	68		FN908096 b
<i>trnG-trnF_M</i> (Spacer)	trnS2 : GAGAGAGAGGGATTGGAACC	trnFm : CATAACCTTGAGGTCACGGG	678	58	FN908093	FN908095 b
<i>trnE-trnT</i> (Spacer)	trnD : ACCAATTGAACTACAATCCC	trnTr : CTACCACCTCAGTTAAAAAGGG	732	52	FN908103	FN908104 b
<i>trnG-trnR</i> (Spacer)	trnS : CGCCGCTTTAGTCCACTCA	trnR : ATTGCGTCCAATAGGATTTGAA	353	58	FN908090	FN908092 c
<i>psbB-psbT</i> (Intron)	psbBF : TGCCCTGGTATCGTGTTCATAC	psbBR : CYTGCTTCTTTGTAGTTGGAT	987	52	FN908099	FN908100 d
<i>petD</i> (Intron)	petB : CTATCGTCCRACCGTTACWGAGGCT	petD : CAAAYGGATAYGCAGGTTCCACC	756	58	FN908102	FN908101 d
<i>trnL</i> (Intron)	TrnL1 : CGAAATCGGTAGACGCTACG	TrnL2R : GGGGATAGAGGGGACTTGAAC	516	62	FN908106	FN908105 e
<i>ndhF-rpl32</i> (Spacer)	ndhF : GAAAGGTATKATCCAYGMATATT	rpl32-R : CCAATATCCCTTYTTTTCCAA	375	54	FN908111	FN908113 e
<i>ndhA</i> (Intron)	ndhAx1 : GCYCAATCWATTAGTTATGAAATACC	ndhAx2 : GGTTGACGCCAMARATTTCCA	628	52	FN908115	FN908114 f
<i>rpl32-trnL</i> (Spacer)	trnL(UAG) : CTGCTTCCTTAAGAGCAGCGT	rpl32-F : CAGTTCCAAAAAACGTACTTC	734	52	FN908116	FN908117 f

Ref: a: (Johnson & Soltis, 1994), b: (Demesure *et al.*, 1995), c: (Dumolin-Lapegue *et al.*, 1997), d: (Grivet *et al.*, 2001), e: (Taberlet *et al.*, 1991), f: (Shaw *et al.*, 2005)

1.3.3 Statistical analyses

➤ Chloroplast sequences

The chloroplast genome is assumed to be non-recombinant, and was thus treated as a single haploid locus. Haplotypes were defined as distinct combinations of alleles found at the two loci *matK-trnK* and *trnQ-trnS*. Sequences were concatenated for each individual and global alignment was performed using the MUSCLE algorithm implemented in CODONCODE ALIGNER 3.5.7 (Codon Code Corporation - Dedham, MA, USA). Phylogenetic relationships among haplotypes were assessed by reconstructing the haplotype network with NETWORK 4.516 software (<http://www.fluxus-engineering.com/>) using the median-joining network algorithm (Bandelt *et al.*, 1999).

➤ Chloroplast SSRs

Haplotypes (chlorotypes) were defined as distinct combinations of alleles at the three cpSSRs. For each population, we computed haplotypic diversity (H_{cp}) (Pons & Petit, 1995), the number of chlorotypes (na_{cp}) and rarefied haplotypic richness (H_R) using the CONTRIB 1.02 program (Petit *et al.*, 1998). H_R is the expected number of different chlorotypes found in each population using a standardized sample size (n) fixed as the smallest available number of genotyped individuals, following the formula given by Comps (Comps *et al.*, 2001). In our case, the smallest population contained only two individuals. Although this statistic is not biased by a small sample ($n=2$), we eliminated seven populations containing fewer than four individuals to better account for rare alleles and to improve precision. The variation trend of haplotypic richness was then examined at different levels of population stratification (between subspecies, between regions and within regions) using Kruskal-Wallis's test implemented in XLSTAT 2009 (Addinsoft). This nonparametric 'one-way analysis of variance' enables comparison of independent samples that do not follow a normal distribution. Following Kruskal-Wallis's test, pairwise comparisons were performed using Dunn's multiple comparison test with Bonferroni correction, XLSTAT 2009 (Addinsoft).

To assess differentiation among subspecies and populations within subspecies, and differentiation among regions and populations within regions, we used hierarchical Analysis



of Molecular Variance (AMOVA) implemented in ARLEQUIN version 3.5 (Excoffier *et al.*, 2005) set up for haploid data.

Two measurements of genetic differentiation described by Pons & Petit (1996) were used for the chlorotypes: G_{ST} differentiation accounting for frequencies, and N_{ST} differentiation accounting for genetic distances between chlorotypes (Pons & Petit, 1996). If the contribution of stepwise mutation to genetic differentiation is negligible, the expected G_{ST} and N_{ST} values are equal. If the differentiation follows a phylogeographic pattern, stepwise mutation would result in $N_{ST} > G_{ST}$. Indeed, under a phylogeographic pattern, phylogenetically related chlorotypes are found more often within than among populations (Pons & Petit, 1996). To compare G_{ST} to N_{ST} , we used a simple test procedure implemented in the SPAGEDI program (Hardy & Vekemans, 2002). The test consists in comparing observed N_{ST} with the distribution of N_{ST} obtained with 1000 estimates (pN_{ST}) under random permutation of the chlorotypes in the distance matrix between chlorotypes. In addition to the global test of N_{ST} vs pN_{ST} , we applied the test in pairwise comparison at different levels of population stratification (between regions, between populations within regions) in order to identify at which scale phylogeographic differentiation occurs.

A matrix of pairwise genetic distances was generated from population chlorotype frequencies using Cavalli-Sforza's chord measure (Cavalli-Sforza, 1967) implemented in the GENDIST program [PHYLIP, version 3.6, (Felsenstein, 1993)]. From this matrix, a dendrogram was constructed using the NJ method of Saitou & Nei (1987) with the NEIGHBOR program of PHYLIP. The significance of each node was evaluated by bootstrapping frequencies over chlorotypes with 1000 replications using the SEQBOOT program of PHYLIP. The consensus tree was obtained with the CONSENSE program of PHYLIP with a cut-off set at 50%, and connections with geographic locations were displayed using GENGIS (Parks *et al.*, 2009).

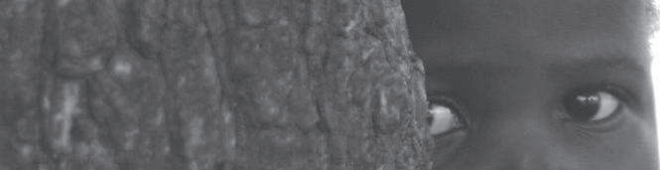
In the same way as for chloroplast sequences, a network was reconstructed using the median-joining network algorithm (Bandelt *et al.*, 1999) implemented in the NETWORK 4.516 program (<http://www.fluxus-engineering.com/>).

➤ Nuclear SSRs

To evaluate polymorphism using nuclear microsatellite markers (nucSSRs), we recorded allele frequencies, number of alleles per locus (na_{nuc}), observed heterozygosity (Ho_{nuc}), and expected heterozygosity (He_{nuc}) using GENETIX 4.05 (Belkhir *et al.*, 2004). Allelic richness (A_R) is defined as the expected number of alleles in each population for a sample size set at the smallest number of individuals genotyped. A_R was calculated with an adaptation of the rarefaction index using FSTAT version 2.9.3.2 (Goudet, 1995). In the same way as for assessment of haplotypic richness, we computed A_R considering only populations with at least four individuals. As for haplotypic richness, variation trends in allelic richness were examined at different levels of population stratification (between subspecies, between regions and within regions) using Kruskal-Wallis's test followed by Dunn's multiple comparison test with Bonferroni correction, implemented in XLSTAT 2009 (Addinsoft).

As for cpSSRs, nuclear-based genetic structure was assessed using ARLEQUIN version 3.5 (Excoffier *et al.*, 2005) and AMOVA. We tested for a phylogeographic signal using the SPAGEDI program (Hardy & Vekemans, 2002), evaluating the contribution of stepwise mutation in the differentiation pattern by comparing observed R_{ST} with R_{ST} obtained after 1,000 allele size permutations (pR_{ST}). If stepwise mutations do not contribute to differentiation, F_{ST} and R_{ST} values are equal, and R_{ST} is expected to be significantly higher than mean permuted R_{ST} under a phylogeographic pattern. In the same way as for cpSSRs, this permutation test was applied to global R_{ST} vs pR_{ST} and in pairwise R_{ST} vs pR_{ST} comparisons at different levels of population stratification (between regions and between populations within regions).

To assess the overall structure of nucSSRs diversity, we used the STRUCTURE program version 2.3.3 (Pritchard *et al.*, 2000). We used an admixture model with correlated allele frequencies, performing 10 independent repetitions for each K (ranging from 1 to 20), with 100,000 Markov chain Monte Carlo repetitions and an 80,000 burn-in period. To identify the optimal K value, we examined absolute values of the second order rate of change of the likelihood distribution divided by the standard deviation of likelihoods (ΔK) as described by Evanno (Evanno *et al.*, 2005). To explore divergence between inferred clusters, we examined the genetic differentiation (F_{ST}) between these clusters and their ancestral population



predicted by STRUCTURE (Pritchard *et al.*, 2000). Phylogenetic relationships among populations were assessed using population allele frequencies with the PHYLIP package (Felsenstein, 1993) using a similar protocol as for cpSSRs.

➤ Bioclimatic data and ecological niche modelling

The ecological niche modelling (ENM) implemented in DIVA-GIS software version 7.2 (Hijmans *et al.*, 2001), enables prediction of the spatial distributions of a species on a bioclimatic basis using the BIOCLIM algorithm (Busby, 1991). From the climate characteristics of given species occurrences, this algorithm gives a percentile distribution (between 0 and 100 percentile) for each climatic variable. Based on homocline matching (Booth *et al.*, 1987), the BIOCLIM algorithm then compares the climatic characteristics of each grid cell in a geographic layer with the percentile distributions of variables inferred from species occurrences. The program, run with the BIOCLIM “classic” option (http://www.diva-gis.org/docs/DIVA-GIS5_manual.pdf), determined six types of area according to their suitability: “not suitable” i.e. one or more climate variables are outside the 0-100 percentile envelope; “excellent” i.e. within the 20-80 percentile envelope for each variable; “very high” i.e. within the 10-90 percentile envelope; “high” i.e. within the 5-95 percentile envelope; “medium” i.e. within the 2.5-97.5 percentile envelope, and marginal areas of “low” suitability for which all variables are at the boundaries of the 0-100 envelope (i.e. between 0-2.5% and 97.5-100%). Considering a 5% threshold, we did not include marginal areas of “low” suitability. We used the occurrence data of *Vitellaria paradoxa* from 126 sample sites (see **Annexe II.2**) and considered seven climatic variables at spatial resolutions of 30 arc-second (about 1 x 1 km): (1) annual mean temperature; (2) mean diurnal temperature range; (3) maximum temperature of the coldest month; (4) minimum temperature of the coldest month; (5) annual precipitation; (6) precipitation seasonality; and (7) precipitation of the wettest quarter. Percentile distributions of these variables were created from current climate conditions and projected onto the layers of current, Last Glacial Maximum (~21,000 years BP), and the last interglacial period (~120,000 - 140,000 years BP). Using current climate conditions (1950–2000) derived from the WorldClim global climate database (Hijmans *et al.*, 2005), ENM enabled us to check that our 126 sample sites were suitable to recover the known actual distribution of *Vitellaria paradoxa* (Hall *et al.* 1996) and thus to describe its climatic preferences. Using data on the Last Glacial Maximum obtained



from the Paleoclimate Modelling Intercomparison Project Phase II (PMIP2) database and achieved under the Community Climate System Model (CCSM3) (Collins *et al.*, 2006), we illustrated the distribution shift of *Vitellaria paradoxa* during this period of major climate perturbations (Maley, 1996). Finally, to describe the potential distribution of the species before LGM perturbations, we predicted the suitability for *Vitellaria paradoxa* using data on the last interglacial (LIG) period generated by Otto-Bliesner (Otto-Bliesner *et al.*, 2006) and provided by the WorldClim global climate database. The performance of the model under current conditions was tested by measuring the area under a relative operating characteristic curve (AUC) (Pearce & Ferrier, 2000) using DIVA-GIS software (Hijmans *et al.*, 2001). AUC is commonly used to evaluate ENM models by measuring their ability to distinguish locations where the species is present from those where it is absent (Hanley & McNeil, 1982). ENM with AUC > 0.7 is considered to perform well (Fielding & Bell, 1997).

1.4 Results

1.4.1 Chloroplast sequences

From alignments of the two polymorphic chloroplast sequences *trnQ-trnS* (601 bp) and *matK-trnK* (452 bp) intergenic spacers, we identified only three single nucleotide polymorphisms (SNPs) among the 40 populations (58 individuals) of *Vitellaria paradoxa*. The combination of these mutations (one located in *trnQ-trnS* and two in *matK-trnK*) resulted in a total of four haplotypes: H1, H2, H3 and H4. H1 was the predominant haplotype with a global frequency of 74.1%, and was shared by West African and Ugandan populations. H2, with a global frequency of 10.3%, was found in Cameroon. H3, found in 5.2% of the individuals, was limited to populations in Chad. Finally, H4 with a global frequency of 10.3%, was found in most individuals in Central African Republic (one tree in southern Central African Republic displayed the H1 haplotype). Therefore, most of the populations were fixed for a single haplotype. We identified a single haplotype among the five individuals of our out-group taxon *Baillonella toxisperma*, a tropical species belonging to Sapotaceae family.

The haplotype network (**Figure II.1**) indicates that haplotypes of *Vitellaria paradoxa* are poorly divergent, and suggests that H2, H3 and H4 derived from the H1 haplotype via a single mutation. The haplotype of the outgroup species was phylogenetically close to *Vitellaria paradoxa* haplotypes, and H1 was closer to the outgroup haplotype than H2, H3

and H4. “H1” populations were located in both West Africa and Uganda, suggesting either a West African or East African origin for existing populations of *Vitellaria paradoxa*.

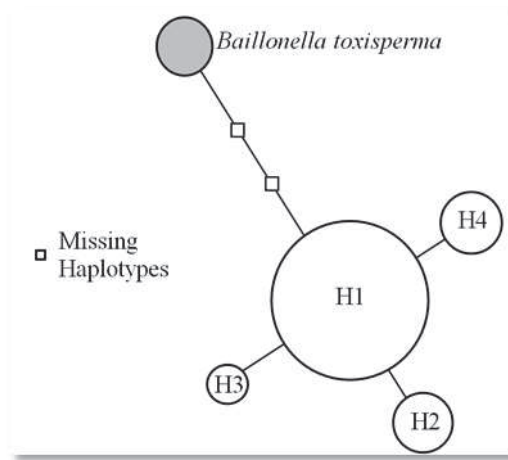


Figure II.1. Phylogenetic relationships among haplotypes (H1, H2, H3 and H4. *Baillonella toxisperma* used as outgroup species) using the median-joining haplotype network (NETWORK 4.516 program).

1.4.2 Chloroplast SSR Markers

➤ Diversity patterns

We identified three polymorphic loci, *ccmp3*, *ccmp5* and *ntcp9*, including respectively three, two and seven variant alleles. These variants were combined into 15 chlorotypes resulting in high total haplotypic diversity with $H_{cp} = 0.9$. However, **Table I.3** shows very poor haplotypic diversity at the population level, with an average of $\hat{H}_{cp} = 0.17$, due to the small population size. **Figure II.2a** shows variation in haplotypic richness (H_R) on a geographic basis. We observed major variations in haplotypic richness H_R among populations, from 0 to 2. No significant differences were detected between subspecies ($P = 0.079$), or between regions ($P = 0.054$), in H_R values using the Kruskal-Wallis test. However, in West Africa, we observed a central zone (around Togo) exhibiting high haplotypic richness values and bounded by two zones with lower H_R (**Figure II.2a**). With the Kruskal-Wallis test, we detected significant differences in haplotypic richness between these areas ($P < 0.001$) and Dunn's pairwise comparisons with Bonferroni correction showed that the central zone around Togo had significantly higher haplotypic richness ($H_R = 1.12$) than the eastern ($H_R = 0.44$) and western ($H_R = 0.09$) flanking zones (Bonferroni-corrected $P = 0.017$) (**Figure II.2a**).

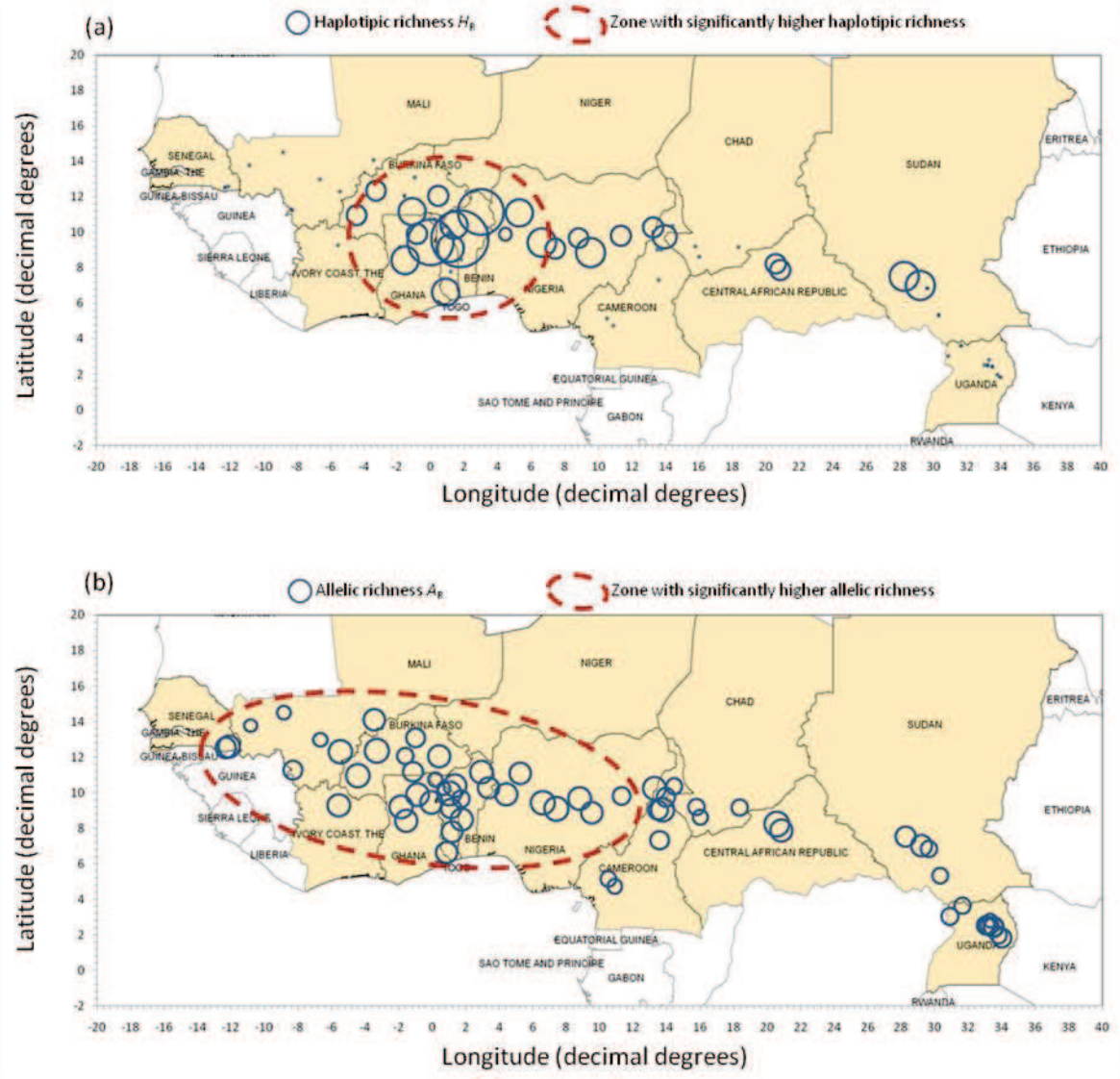


Figure II.2. Trends in haplotypic (cpSSRs) and allelic (nucSSRs) richness. Diameters of circles vary proportionally with richness values. Areas bounded by a red dotted line had significantly higher values than others with Dunn's comparison test with Bonferroni correction. (a) Trends in haplotypic richness H_R . (b) Trends in allelic richness A_R .

➤ Differentiation

We identified 15 chlorotypes in the natural area (**Figure II.3**) with frequencies ranging from 1.1% (D) to 19.8% (C). Eleven chlorotypes were found in the subspecies *paradoxa*, including eight haplotypes (B-C-F-G-H-L-N-O) limited to West African populations, and three (A-I-M) restricted to Central Africa. Two private haplotypes (E-K) were limited to *nilotica* subspecies. Two haplotypes (D-J) were shared between the subspecies in Central African Republic (*paradoxa*) and Sudan (*nilotica*). Evaluation of differentiation using hierarchical AMOVA revealed populations to be significantly differentiated at subspecies level, with 37% of the variation detected among subspecies ($F_{CT} = 0.37$, $P = 0.00$), and 57% among

populations within subspecies ($F_{SC} = 0.91$, $P = 0.00$), and only 6% of variation within populations. Differentiation at regional level (West Africa, Central Africa and East Africa) was even stronger with 58% of genetic variation detected among geographical groups ($F_{CT} = 0.58$, $P = 0.00$), 37% among populations within groups ($F_{SC} = 0.87$, $P = 0.00$), and only 5% within populations.

Table II.3. Diversity parameters calculated per population and averaged per country. Ns: number of populations; N: number of individuals; Nacp: number of haplotypes per country; Hcp: haplotypic diversity; Necp: number of effective haplotypes per population; Nanuc: mean number of alleles per population and per locus; Honuc: observed heterozygosity; Henuc: expected heterozygosity on Hardy-Weinberg equilibrium.

Region	Country	Ns	N	Chloroplast microsatellites				Nuclear microsatellites				Subsp. Species
				N_{cp}	H_{cp} (s.d)	$N_{e_{cp}}$ (s.d)		N_{nuc}	H_{nuc} (s.d)	$H_{e_{nuc}}$ (s.d)		
West Africa	Senegal	2	9	1 L	0.00 (0.00)	1.00 (0.00)		3.42	0.53 (0.26)	0.55 (0.20)		ssp. <i>Paradoxa</i>
	Mali	7	29	2 CL	0.00 (0.00)	1.00 (0.00)		2.68	0.47 (0.33)	0.43 (0.23)		ssp. <i>Paradoxa</i>
	Ivory Coast	2	6	1 C	0.00 (0.00)	1.00 (0.00)		3.00	0.67 (0.26)	0.54 (0.18)		ssp. <i>Paradoxa</i>
	Burkina-Faso	6	36	3 BCF	0.24 (0.21)	1.11 (0.27)		3.49	0.59 (0.30)	0.55 (0.23)		ssp. <i>Paradoxa</i>
	Ghana	4	24	3 BCN	0.31 (0.22)	1.42 (0.50)		4.23	0.60 (0.26)	0.59 (0.21)		ssp. <i>Paradoxa</i>
	Togo	6	32	4 CFGH	0.15 (0.24)	1.11 (0.27)		3.28	0.54 (0.33)	0.48 (0.23)		ssp. <i>Paradoxa</i>
	Benin-Niger	6	30	4 CFGH	0.28 (0.32)	1.11 (0.27)		3.49	0.55 (0.36)	0.51 (0.25)		ssp. <i>Paradoxa</i>
	Nigeria	8	65	4 GMNO	0.27 (0.09)	1.71 (0.12)		4.57	0.56 (0.27)	0.57 (0.22)		ssp. <i>Paradoxa</i>
	Cameroon	7	35	2 AM	0.06 (0.15)	1.14 (0.38)		2.74	0.53 (0.32)	0.48 (0.20)		ssp. <i>Paradoxa</i>
	Chad	3	14	1 A	0.00 (0.00)	1.00 (0.00)		2.28	0.55 (0.37)	0.42 (0.22)		ssp. <i>Paradoxa</i>
Central Africa	C. Afr. Rep.	3	15	3 DIJ	0.22 (0.19)	1.33 (0.58)		3.86	0.59 (0.25)	0.62 (0.14)		ssp. <i>Paradoxa</i>
East Africa	Sudan	7	26	4 DEJK	0.24 (0.31)	1.24 (0.42)		2.49	0.52 (0.30)	0.46 (0.18)		ssp. <i>Nilotica</i>
	Uganda	10	43	1 E	0.00 (0.00)	1.00 (0.00)		2.60	0.54 (0.31)	0.47 (0.20)		ssp. <i>Nilotica</i>

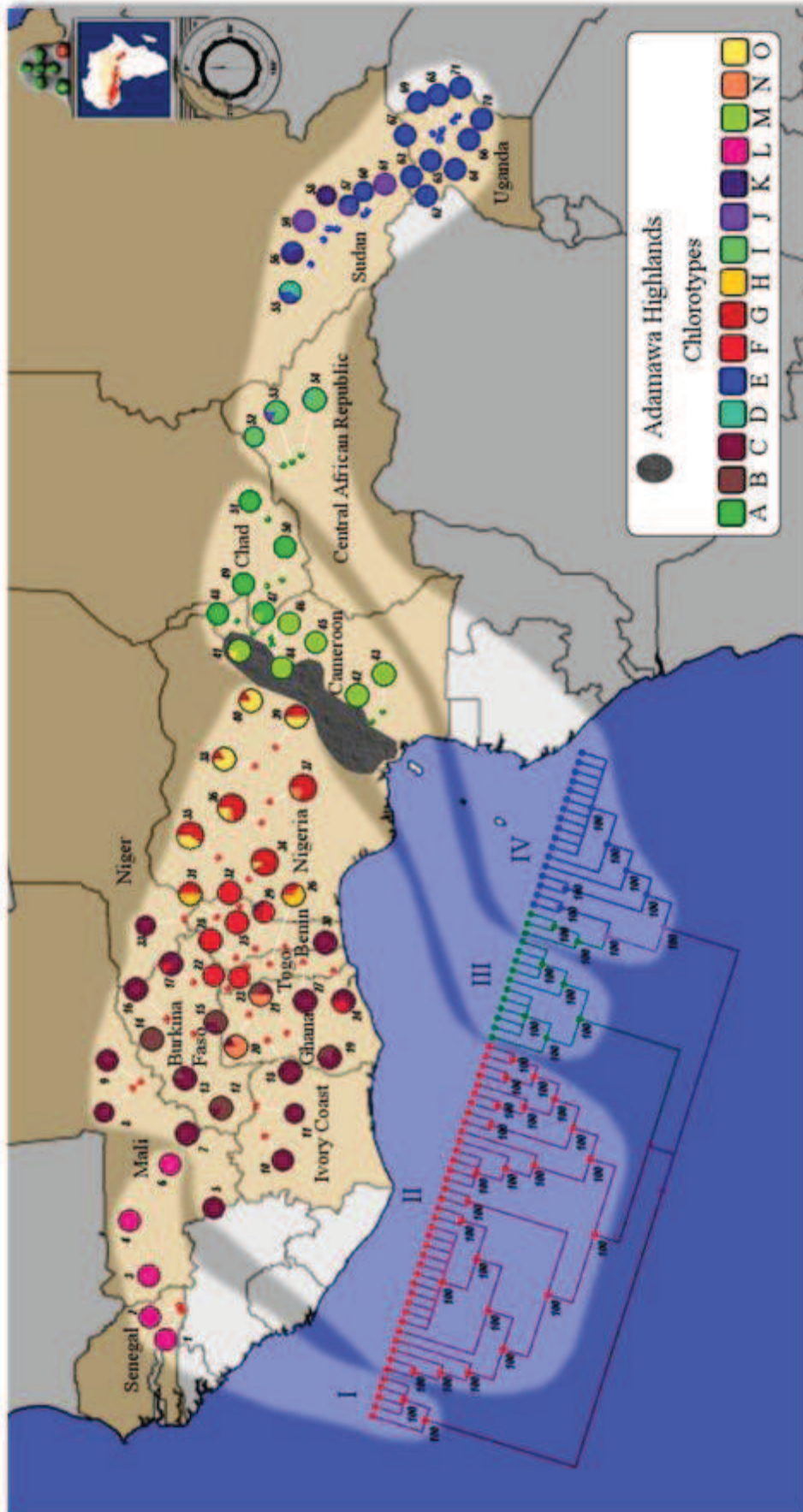


Figure II.3. Geographical distribution of the 15 chlorotypes of *V. paradoxa*. Unrooted neighbor-joining tree drawn with GenGIS software (Parks et al., 2009) using Cavalli-Sforza genetic distances. Numbers at the base of the branches are percentages corresponding to the bootstrap values after 1 000 replications.

1: Bandafassi; 2: Samecuta; 3: Mahina; 4: Torodo; 5: Badougou; 6: Kodougouni; 7: Mperesso; 8: Songo; 9: Tenndeli; 10: Napieo ledougou; 11: Korhogo; 12: Peni; 13: Karo; 14: Bonogo; 15: Pô; 16: Konia; 17: Fada N'Gourma; 18: Larabanga; 19: Kawampe; 20: Kadia; 21: Yendi; 22: Tabango; 23: Payô; 24: Koudassi; 25: Kante; 26: Kolina; 27: Djangbassou; 28: Cocota; 29: Kpale; 30: Pira; 31: Kandi; 32: Kalale; 33: Boumba; 34: Wawa; 35: Rijau; 36: Paiko; 37: Abuja; 38: Kuru; 39: Shendam; 40: Kaltungo; 41: Mubi; 42: Bangangte; 43: Foumban; 44: Ngong; 45: Ngaoundere; 46: Mafa-Kilda; 47: Figuil; 48: Mendeo; 49: Kelo; 50: Mondou; 51: Sarh; 52: North; 53: Central; 54: South; 55: Kuajena; 56: Cuebet; 57: Wulu; 58: Kuel Kuac; 59: Matengai; 60: Mbara; 61: Mundri; 62: Arua; 63: Eria; 64: Adumara; 65: Otongere; 66: Okwongo; 67: Lacankweri; 68: Olwoko; 69: Otuks; 70: Abarilela; 71: Akoboi

➤ Phylogeographic patterns

Overall differentiation among populations was very high and significant with $G_{ST} = 0.81$ ($P = 0.00$) based on unordered alleles, and $N_{ST} = 0.90$ ($P = 0.00$) based on ordered alleles. The global test of phylogeographic structure was significant ($N_{ST} > pN_{ST}$, $P < 0.001$) with $pN_{ST} = 0.80$, revealing a clear phylogeographical pattern in chloroplast variation. However, this phylogeographical signature was no longer detectable within regions (West: $P = 0.18$; Central: $P = 0.60$; and East: $P = 0.49$) indicating that mutation did not contribute to chloroplast differentiation within geographical zones. Pairwise comparisons of N_{ST} vs pN_{ST} between West and East Africa revealed a highly significant phylogeographic signal ($N_{ST} = 0.69$ and $pN_{ST} = 0.35$ with $P = 0.00$), but not between West and Central Africa ($N_{ST} = 0.36$ and $pN_{ST} = 0.23$ with $P = 0.13$), or between Central and East Africa ($N_{ST} = 0.38$ and $pN_{ST} = 0.4$ with $P = 0.47$).

Phylogenetic relationships between populations were supported by strong bootstrap values (**Figure II.3**). The frequency-based NJ tree displayed four clades (I-IV) which perfectly match the geography. Clade I contains marginal western populations (populations 1 to 4 and 6) suggesting their isolation from other West African populations (populations 5 and 7 to 42). Clade II contains other West African populations. Clade III includes most Central African sites, apart from Central African Republic. Clade IV contains eastern populations, combining the subspecies *paradoxa* (Central African Republic) and *nilotica* (Sudan and Uganda).

Phylogenetic relationships among chlorotypes illustrated in the median-joining network (**Figure II.4**) suggested global differentiation following a phylogeographic pattern, but with some inconsistent reticulations. Private West African chlorotypes were closely related, defining group 1 in the network. Within this group, each chlorotype was separated from two to four others by only 1 bp, explaining the lack of a phylogeographic pattern in West Africa. Group 2 in the network includes the three East African *nilotica* chlorotypes. These East African chlorotypes were connected to West African chlorotypes: the network showed that L (found in Senegal) and K (found in Sudan) were separated by only 1 bp. Central African chlorotypes were split into two distant groups in the network. Haplotype M (group 3) was the only chlorotype from southern Cameroon related to “N”, a West African chlorotype found in Nigeria. However, the four missing chlorotypes between M and N

Figure 1 is a network diagram illustrating the relationships between 12 haplotypes (A-L) and their evolutionary distances. The diagram is divided into four numbered regions (1, 2, 3, 4) by dashed lines. Region 1 contains haplotypes B, C, F, G, H, and L. Region 2 contains haplotypes E, J, and K. Region 3 contains haplotypes M and N. Region 4 contains haplotypes A, D, and I. Haplotypes are represented by colored circles of varying sizes, and missing haplotypes are represented by small squares. The connections between haplotypes represent evolutionary distances, with some connections marked by vertical bars indicating specific distances.

1.4.3 Nuclear SSR Markers

A total of 173 alleles were detected at 12 nuclear SSR loci in our collection, with an average of 14.41 alleles per locus. Low variation was observed in mCIRVp08 and mCIRVp175 with five alleles, while mCIRVp167 was highly variable with 25 alleles. Expected and observed (in parentheses) heterozygosity varied markedly from $He_{nuc} = 0.27$ ($Ho_{nuc} = 0.26$) in Mahina (Mali) to $He_{nuc} = 0.65$ ($Ho_{nuc} = 0.74$) in Kanté (Togo). Average values per country are listed in **Table I.3**.

Allelic richness (A_R) varied from 1.8 to 3.4 among populations, but with no clear trend (**Figure II.2b**). However, the Kruskal-Wallis test revealed significant differences between *paradoxa* ($A_R = 2.8$) and *nilotica* ($A_R = 2.4$) **subspecies** ($P = 0.016$) with significantly higher A_R for *paradoxa* with Dunn's pairwise comparison test with Bonferroni correction (Bonferroni-corrected $P = 0.05$). In addition, the Kruskal-Wallis test detected significant differences in allelic richness between geographical regions ($P = 0.006$) and Dunn's pairwise comparisons with Bonferroni correction highlighted significantly higher allelic richness in West Africa ($A_R = 2.9$) than in Central ($A_R = 2.5$) and Eastern ($A_R = 2.4$) Africa (Bonferroni-corrected $P = 0.017$) (**Figure II.2b**).

➤ Differentiation

In the same way as with chloroplast SSRs, with nuclear SSR data, AMOVA revealed that the subspecies *nilotica* and *paradoxa* were significantly differentiated: 22% of variation was detected between subspecies ($F_{CT} = 0.22$, $P = 0.00$), and 17% between populations within subspecies ($F_{SC} = 0.21$, $P = 0.00$). Likewise, differentiation was significant at the scale of geographic regions, with 19% of variance explained by the regions ($F_{CT} = 0.19$, $P = 0.00$) and 16% among populations within regions ($F_{SC} = 0.19$, $P = 0.00$). We can thus assume that geographic distribution is a strong determinant of the nuclear-based genetic structure of *V. paradoxa*. However, up to 65% of the variation was detected within populations, in contrast to 5% with chloroplasts.

Using the Bayesian-based approach of Pritchard (Pritchard *et al.*, 2000), two peaks of ΔK for $K = 2$ and $K = 4$ were detected according to the method of Evanno (Evanno *et al.*, 2005). The clustering obtained for $K = 2$ (**Figure II.5a**) separated populations into two geographically consistent groups with low levels of admixed individuals: the "West" group included populations from West Africa and the "East" group included the remaining populations. This clustering in two groups constituted the upper level of structure in *Vitellaria paradoxa*. Although $K = 2$ was identified as the best model, $K=4$ (**Figure II.5b**) obtained high probability and revealed very clear substructures in higher groups. Within the "East" group, we detected a substructure that distinguished a "Central" (C, ssp. *paradoxa*) from an "Eastern" group (E, ssp. *nilotica*), consistent with the distribution of the two subspecies in this zone. In West Africa, populations were also split in two groups consistent

with their geographical distribution, “West 1” (W1) including most western populations and “West 2” (W2) populations from Nigeria. However, W1 included a large number of admixed individuals. Bayesian analysis revealed migrants and admixed individuals between the “Central” and “Eastern” clusters of STRUCTURE, and thus between *nilotica* and *paradoxa* (**Figure II.5b and II.6**). In comparison to the ancestral population predicted by STRUCTURE, clusters W1 ($F_{ST} = 0.06$) and W2 ($F_{ST} = 0.05$) displayed the least differentiation, whereas clusters C ($F_{ST} = 0.12$) and E ($F_{ST} = 0.23$) diverged further from the potential ancestral population. A PCA analysis on nuclear microsatellite allele frequencies was highly congruent with the STRUCTURE analysis (results not shown).

➤ Phylogeographic patterns

Overall differentiation among populations assessed with nuclear SSRs was high and significant for both the index based on unordered alleles ($F_{ST} = 0.21$, $P = 0.00$) and that based on ordered alleles ($R_{ST} = 0.28$, $P = 0.00$). Comparisons of R_{ST} and pR_{ST} revealed that the nuclear microsatellite diversity pattern of *Vitellaria paradoxa* had a significant phylogeographic signature with $pR_{ST} = 0.21$ and $R_{ST} > pR_{ST}$ ($P = 0.004$). The phylogeographic signal was highly significant for each pairwise regional comparison: between West and Central Africa ($R_{ST} = 0.16$ and $pR_{ST} = 0.07$ with $P = 0.005$), between West and East Africa ($R_{ST} = 0.27$ and $pR_{ST} = 0.07$ with $P = 0.00$) and between Central and East Africa ($R_{ST} = 0.25$ and $pR_{ST} = 0.08$ with $P = 0.00$). However, as for chloroplasts, variation trends within geographic regions did not match the phylogeographic pattern (West: $P = 0.49$; Central: $P = 0.25$; East: $P = 0.34$), suggesting a major contribution of migration drift within geographical regions. Finally, we performed pairwise comparison between cluster groups “W1” and “W2” inferred by STRUCTURE (**Figure II.4b**) but detected no significant phylogeographic signal ($R_{ST} = 0.06$ and $pR_{ST} = 0.04$ with $P = 0.11$).

The degree of admixture of each population linked to their geographical locations confirmed that the clustering obtained with the Bayesian approach fitted the geographical distribution (**Figure II.6**). Moreover, higher levels of admixture were found at geographical junctions between clusters. Within the W1 group, admixed populations were most often found in Togo, Benin and Niger (populations 24, 25, 28, 29, 30, 31, 32 and 33). Populations 52, 53, 54, 56 and 60, at the boundary between *paradoxa* and *nilotica* zones were also

admixed (Figure II.6). Phylogenetic relationships at population level based on allele frequencies were poorly informative as nodes were not significant (data not shown).

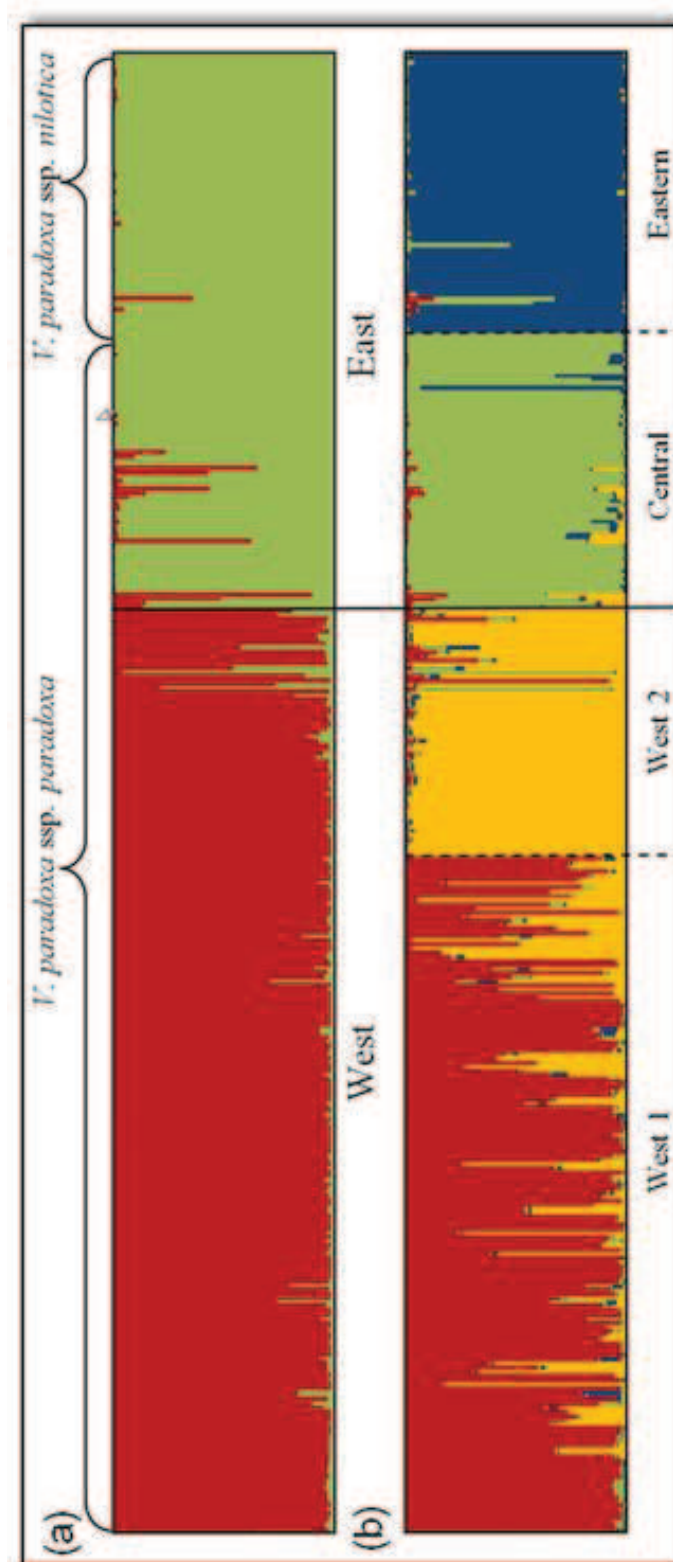


Figure II.5. (a) Bar plot showing clustering of individuals by STRUCTURE with $K = 2$ (Pritchard *et al.*, 2000). (b) Bar plot showing clustering of individuals by STRUCTURE with $K = 4$.

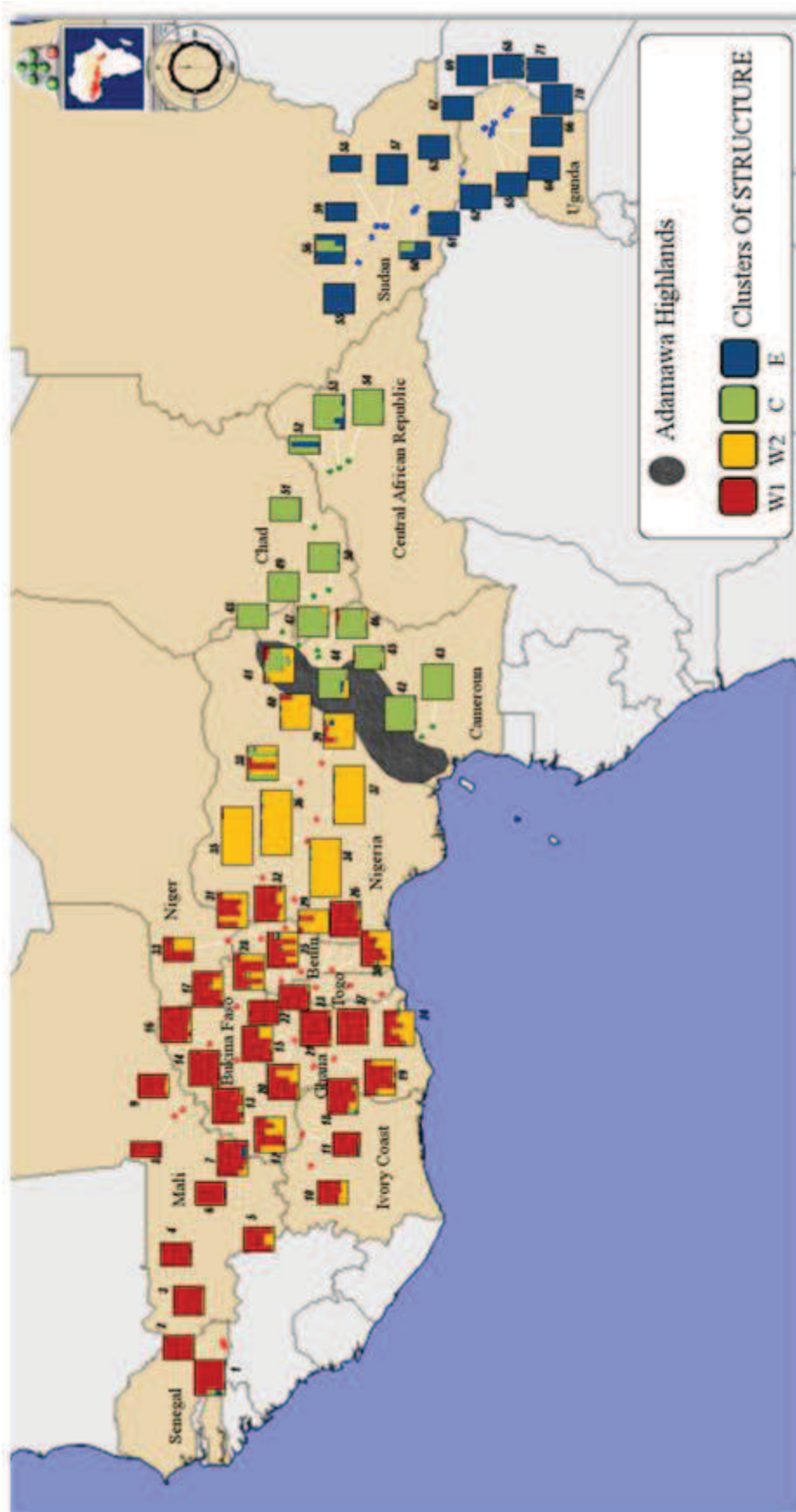
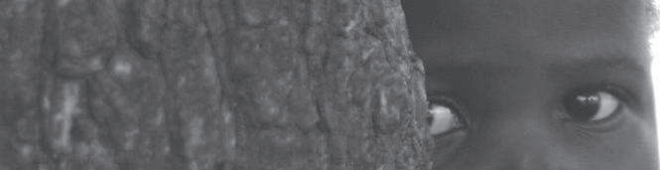


Figure II.6. Geographical plot of clustering of individuals by structure with $K = 4$: W1: "West 1", W2: "West 2", C: "Central", E: "Eastern".

1: Bandafassi; 2: Samecoua; 3: Mahina; 4: Torodo; 5: Badougou; 6: Kodougouni; 7: Mperesso; 8: Songo; 9: Tenneli; 10: Napieledougou; 11: Korhogo; 12: Peni; 13: Karo; 14: Bonogo; 15: Pô; 16: Konia; 17: Fada N'Gourma; 18: Larabanga; 19: Kawampe; 20: Kadia; 21: Yendi; 22: Tabango; 23: Payô; 24: Koudassi; 25: Kante; 26: Kolina; 27: Djangbassou; 28: Cocota; 29: Kpale; 30: Pira; 31: Kandi; 32: Kalale; 33: Boumba; 34: Wawa; 35: Rijau; 36: Paiko; 37: Abuja; 38: Kuru; 39: Shendam; 40: Kaltungo; 41: Mubi; 42: Bångangte; 43: Fouban; 44: Ngong; 45: Ngaoundere; 46: Mafa-Kilda; 47: Figuil; 48: Mendeo; 49: Kelo; 50: Mondou; 51: Sarh; 52: North; 53: Central; 54: South; 55: Kuajena; 56: Cuebet; 57: Wulu; 58: Kuel Kuac; 59: Matengai; 60: Mbarai; 61: Mundri; 62: Arua; 63: Eria; 64: Adumara; 65: Otongere; 66: Okwongo; 67: Lacankweri; 68: Olwoko; 69: Otuks; 70: Abarilela; 71: Akoboi



1.4.4 Present and previous theoretical distribution of *Vitellaria paradoxa*

From our database containing 126 sample sites, we checked that the zone theoretically suitable for the species under current conditions (**Figure II.7a**) was consistent with its known distribution (Hall *et al.* 1996). A decrease in suitability was observed at the junctions between the geographic regions (West, Central and East) suggesting lower species density, and extending areas suitable for *Vitellaria paradoxa* to regions where the species is in fact absent (below the equator). The AUC, averaged over the seven climatic variables, was scored at 0.79, indicating that the accuracy of the model was reasonably good for ENM predictions. Ecological niche modelling of the species during the Last Glacial Maximum (~21000 years BP) showed a much smaller suitable zone (**Figure II.7b**): while West Africa remained highly suitable, the putative distribution of *Vitellaria paradoxa* from Cameroon to Uganda was reduced and fragmented. In this part of Africa, three putative areas were detected where the species could have been limited: near Lake Chad, in southern Central African Republic, and in northern Uganda (see zones 1, 2 and 3 in **Figure II.7b**). Finally, we predicted probable suitable areas for the Shea tree during the last interglacial period around 120000-140000 BP (**Figure II.7c**). Results showed a potential distribution pattern divided into two main zones. The first zone extended from present-day Senegal to the north of Central African Republic, but the zone was almost completely disrupted at the Nigerian border. The second zone area mainly comprised Sudan and northern Uganda. These zones were separated by a large gap (~200 km) at the border between Sudan and Central African Republic, consistent with the current geographical gap between the two subspecies *nilotica* and *paradoxa* (Hall *et al.*, 1996).

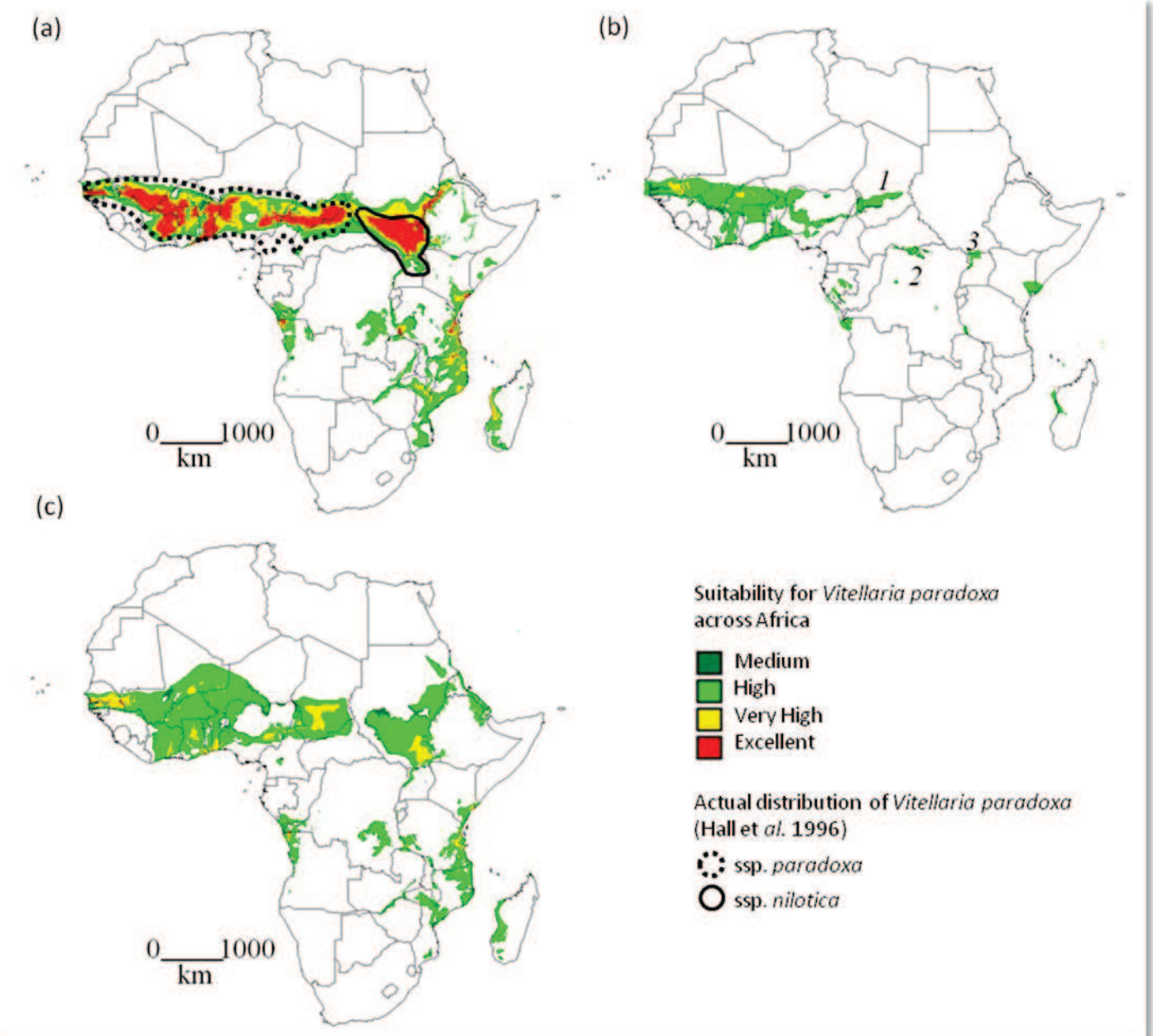
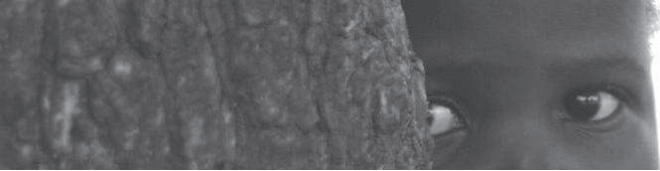


Figure II.7. Ecological niche modelling of *Vitellaria paradoxa* at three different periods: (a) today predicted and reported natural range; (b) potential distribution during the Last Glacial Maximum (20000 years BP) and refugia (1: around Lake Chad; 2: northern Congo; 3: northern Uganda); (c) potential natural range during the last interglacial period (~120000-140000 years BP). Colours represent the probability of suitable climatic conditions from red (excellent) to green (medium). Envelopes were produced using the Worldclim/PMIPII data base and diva-gis software version 7.2 (Hijmans et al., 2001).

1.5 Discussion

1.5.1 Genetic diversity of *Vitellaria paradoxa*

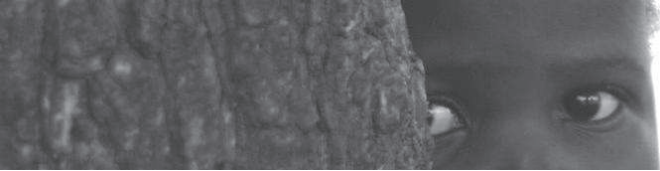
In our study, nuclear and chloroplast SSRs genetic diversity was higher than that observed in previous studies on *Vitellaria paradoxa* (Fontaine et al., 2004; Sanou et al.,



2005). This novel assessment based on a larger sample is consistent with diversity detected in angiosperms in other studies (Müller *et al.*, 2009) and meets expectations concerning the diversity of widespread tree species with similar biological attributes (Hamrick *et al.*, 1993). However, we noted that the level of SSR diversity varied markedly among zones in our study (**Figure II.2**). Possible explanations for higher levels of diversity are the presence of ancient glacial refugia or the confluence of migration routes (Petit *et al.*, 2003) and we propose some hypotheses below. In addition to SSR markers, polymorphism within chloroplast intergenic spacers revealed very weak diversity with only four haplotypes. This poor variability was comparable to the diversity detected with chloroplasts in other widespread tree species such as *Pinus pinea* L. (Vendramin *et al.*, 2008) and *Adansonia digitata* (baobab), another widespread savanna tree (Pock Tsy *et al.*, 2009). Therefore, in contrast with species harbouring high levels of polymorphism of intergenic spacers (Dick & Heuertz, 2008), these markers are not an effective tool to study the phylogeography of *Vitellaria paradoxa*. However, the low polymorphism detected in these sequences was consistent with the low rate of nucleotide substitution in chloroplast genomes, around 10^{-9} (Wolfe *et al.*, 1987), compared to SSR mutation rates of from 10^{-4} to 10^{-2} (Vigouroux *et al.*, 2002) with nuclear SSRs and from 10^{-5} to 10^{-3} with chloroplast SSRs (Navascués & Emerson, 2005).

1.5.2 A phylogeographic pattern driven by climate changes

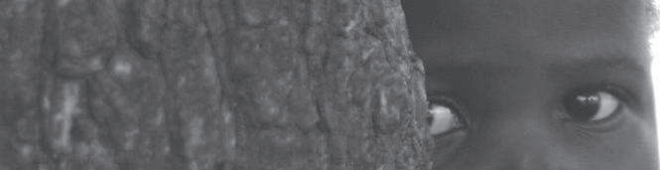
We showed that the genetic diversity of *Vitellaria paradoxa* has a clear east-west structure (**Figures II.3, II.4, II.5 and II.6**) and chloroplasts revealed a strong phylogeographic signature ($G_{ST} = 0.81$; $N_{ST} = 0.90$; $N_{ST} > pN_{ST}$) and nuclear SSRs ($F_{ST} = 0.21$; $R_{ST} = 0.28$; $R_{ST} > pR_{ST}$). We identified two major genetic groups geographically separated by the Adamawa Highlands: a “West” group, which is relatively homogeneous over the 3000 km area comprising West Africa, and an “East” group divided into two subgroups (**Figures II.3 and II.4**). We propose that the strong phylogeographic pattern of *Vitellaria paradoxa* is explained by past climate shifts and that the last glacial maximum (LGM) perturbation is a major factor in explaining this pattern. During the LGM, the climate in Africa was colder and drier; this shift was particularly marked in eastern sub-Saharan Africa and was characterized by small incursions of the Sahara desert into this region (Lioubimtseva *et al.*, 1998).



Consistent with these climatic perturbations, ENM predictions indicated that the potential distribution of *Vitellaria paradoxa* in this zone was highly fragmented (**Figure II.7b**). However, ENM results should be interpreted with caution because they depend on the reliability of the climatic models (current and ancient) resulting from mathematical interpolations and global smoothing of collected data (Hijmans *et al.*, 2005). For example, the poor quality of climatic documentation resulted in local over-estimation of the potential distribution of a mountain bird species (*Cranioleuca henricae*) in Bolivia (Soria-Auza *et al.*, 2010). Consequently, we suggest only using ENM to describe general trends of species range in our study. In addition, ENM predictions assume a no evolution model and thus do not consider the adaptation ability and the dispersal dynamics of species across ages. Having taken these limitations into account, using ENM on LGM data, we identified two reduced areas which could have formed glacial refugia for populations of *Vitellaria paradoxa* in this part of Africa: around Lake Chad and in northern Uganda, along with other marginal refugia (**Figure II.7b**). We suggest that these refugia could be at the origin of the genetic divergence of *Vitellaria paradoxa* in eastern sub-Saharan Africa. In contrast to the eastern area, West Africa was less impacted by the climate change during the LGM, probably due to the effect of the monsoon from the Gulf of Guinea (Braconnot *et al.*, 2000). As a result, while the eastern suitable area for *Vitellaria paradoxa* was probably deeply disturbed, ENM results suggest that distribution of the species was less impacted and thus probably conserved in West Africa (**Figures II.7b and II.7c**).

While the limited LGM perturbation in West Africa explains the high level of diversity observed and the hotspot detected in its central area (around Togo) (**Table I.3 and Figure II.2**), the poor structure detected in this large region (**Figures II.5 and II.6**) could also result from the intense human activity in the area. Palynological studies reported human migrations and agropastoral activities involving Shea trees 1,000 years ago in Burkina (Neumann *et al.*, 1998), and evidence for human selection of *Vitellaria paradoxa* has been found in this zone (Lovett & Haq, 2000; Maranz & Wiesman, 2003).

In addition to past climate variations, we propose that the strong differentiation between West and Central African populations was probably preserved by topographic elevations. For instance, the Adamawa Highlands clearly represent a boundary between West African and other populations. Peaking at 1,300 metres asl. between Nigeria and



Cameroon (**Figures II.3** and **II.4**), this mountain range creates a colder climate resulting in a local barrier to *Vitellaria paradoxa* colonization and limiting gene flow between West and Central Africa. Therefore, apart from the possible secondary role of Adamawa, we propose that the pan-continental phylogeographic pattern of *Vitellaria paradoxa* is fully consistent with perturbations during the last glacial maximum, and that this pattern was little affected by the dispersal dynamics of the species.

1.5.3 Differentiation between *paradoxa* and *nilotica*

Results in the literature argue for species differentiation in two non-overlapping subspecies based on slight variations in their leaf and stamina morphology (Hall *et al.*, 1996). We consequently expected to detect a prior genetic differentiation based on this distinction. However, several molecular results suggest that this taxonomic distinction should be reconsidered. The median-joining network obtained with chloroplast SSRs (**Figure II.4**) revealed that Central African chlorotypes (*paradoxa*) are closer to *nilotica* than to other *paradoxa* chlorotypes. In addition, the nuclear SSRs Bayesian approach revealed a clear and strong differentiation of the species in two groups on the two sides of the Adamawa Highlands: a western group comprising West African populations (ssp. *paradoxa*) and an eastern group including *nilotica* and *paradoxa* from Cameroon to Uganda (**Figure II.5a**). The molecular distinction between subspecies *nilotica* and *paradoxa* appears only as a substructure of the eastern group for $K = 4$ (**Figure II.5b**). These results suggest that *paradoxa* populations in eastern sub-Saharan Africa (from Cameroon to Central African Republic) share a common evolutionary history with *nilotica* rather than with other *paradoxa* from West Africa, tightened by the inter-subspecies migrations and hybridizations revealed in our study (**Figures II.5** and **II.6**). Moreover, migration events between subspecies were also supported by chloroplast microsatellites. Indeed, the *nilotica* migrant detected with nuclear SSRs in Central African Republic shared *nilotica* chlorotype ("J"), and two admixed individuals found in Sudan shared a *paradoxa* chlorotype ("D") (**Figures II.3** and **II.4**). We cannot exclude that these shared chlorotypes are due to homoplasy or shared ancestral polymorphism rather than to migration. However, the chloroplast microsatellite mutation rate is likely to be low, reducing the effect of homoplasy (Navascués & Emerson, 2005). Based on this molecular differentiation, we suggest that the distinction between subspecies perhaps occurred spatially on the two sides of the Adamawa Highlands, which



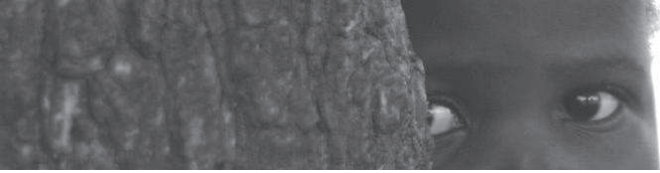
may be consistent with the trend observed in lipid composition suggesting that Shea butter in Cameroon is of the “*nilotica*” type (Maranz *et al.*, 2004).

1.5.4 Hypothesis on the origin of *Vitellaria paradoxa*

While no evidence in the literature proposes an origin for *Vitellaria paradoxa*, combined results obtained with nuclear and chloroplast markers provided new insights into present populations. Using both chloroplast and nuclear SSRs, higher diversity was detected in West Africa, particularly around Togo (**Fig 2, Table I.3**), suggesting ancient populations in this area (Petit *et al.*, 2003). This result has been confirmed by another study (Logossa, personal communication) on 674 individuals from 38 populations sampled throughout West Africa. In addition, Bayesian analysis of nuclear SSRs showed that West African groups displayed minimal divergence from their theoretical ancestor ($F_{ST} = 0.05$), whereas other groups diverged more (Central: $F_{ST} = 0.12$; East: $F_{ST} = 0.23$). This result indicates a historically higher effective population size in West Africa, which could be compatible with the origin of *Vitellaria paradoxa* being in this zone. This hypothesis fits with the recently inferred West African origin of the baobab (Pock Tsy *et al.*, 2009), another African savanna tree whose distribution north of the equator is very similar to that of *Vitellaria paradoxa*. However, low F_{ST} relative to ancestral populations and higher diversity in West Africa do not necessarily result from a more ancient origin of the species compared to the other regions. To be able to conclude on the origin of *Vitellaria paradoxa*, additional DNA sequence data are required since chloroplast sequence analysis indicates the presence of likely ancestral haplotype both in West Africa and Uganda (**Figure II.1**).

1.5.5 Outlook for conservation of *Vitellaria paradoxa* in the African savanna biome

The savanna is demographically, environmentally and economically a major African ecosystem which requires urgent management efforts (Sankaran *et al.*, 2005). Despite its importance, forecasts of anthropogenic climate change on savanna biome are not lacking. In our study, we identified West Africa as a diversity hotspot for *Vitellaria paradoxa*, possibly due to a weaker impact of climate shifts on this savanna tree in this zone. Given the greater past perturbations detected farther east, we suggest that current global warming could



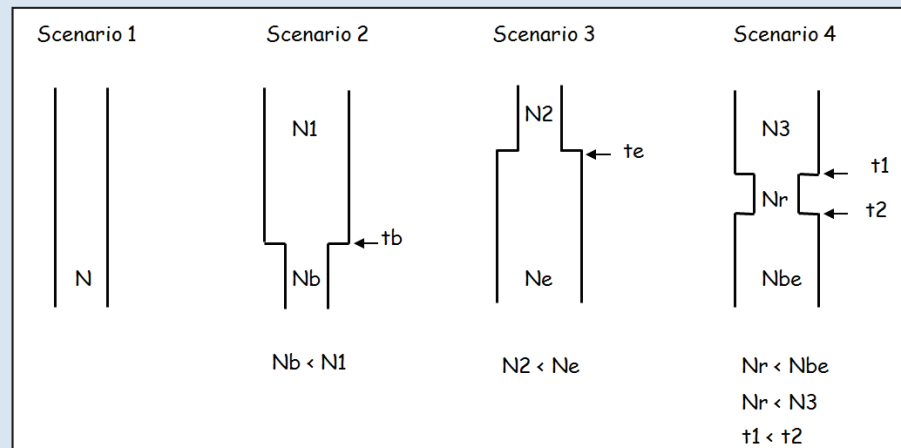
Chapitre II : Étude des Processus historiques et démographiques

affect this part of the species area first, and by extension, the associated biotope. Consequently, major efforts should be made to protect the species in this zone. Due to the possible buffering effect of the monsoon in upper Guinea, West African *Vitellaria paradoxa* populations should be less affected by the current climate shift, but we nevertheless suggest that efforts should be increased to maintain the diversity of this zone, including identifying and protecting the numerous *Vitellaria paradoxa* landraces in West Africa.

BOX II.1 DETECTING POPULATION EXPANSION IN WEST AFRICA

We used the Approximate Bayesian Computation (ABC) approach as implemented in DIYABC (Cornuet *et al.*, 2008; Cornuet *et al.*, 2010) to determine the demographic scenario that may have shaped West African Shea tree diversity pattern. ABC is a flexible class of Monte-Carlo algorithms used to perform model-based inference (Beaumont *et al.*, 2002; Csilléry *et al.*, 2010), using coalescence tool to generate thousands of simulated data points from a set of user-defined demographic scenarios with parameters sampled from biologically meaningful prior distributions and constituting the further called reference table. Choosing simulated points that are close enough to the observed data according to an Euclidian distance criterion and on the basis of summary statistics, it allows to identify the scenario that more likely generated the observed data and to estimate its parameter posterior distributions.

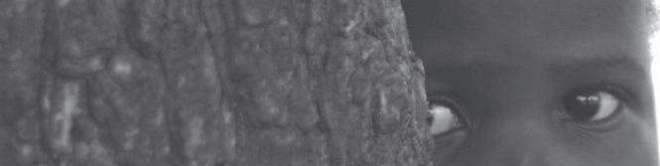
Figure II.8. Scenarios tested using DIYABC (Cornuet *et al.*, 2008)



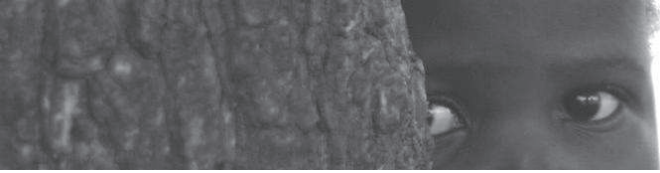
Main Results & Discussion

Consistently with the results presented in phylogeography section, we considered West African Shea trees as a single big panmictic population and proposed 4 equiprobable scenarios (Figure II.8). The prior constraints were chosen as broad as possible with in mind biological and historical consistency. We then built the reference table with about 500000 simulated data points, evenly distributed across scenarios. We recorded mean allele number, mean genetic diversity (Nei, 1987), mean allele size variance and mean Garza-Williamson's M (Garza & Williamson, 2001; Excoffier *et al.*, 2005) as sample statistics to estimate scenario probability, parameter confidence interval and parameter bias and error as well as to assess confidence in scenario choice.

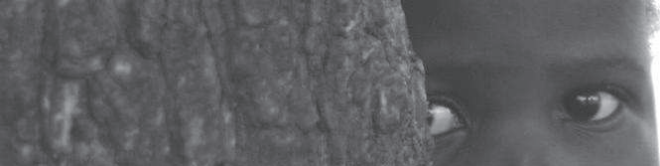
The results of our ABC analysis strongly support a population expansion of Shea tree in West Africa. First, probabilities for the two scenarios involving an expansion event (scenario 3 and 4) are very high and have narrow confidence interval, and second, the alpha risk of detecting an expansion in case of constant population size or bottleneck with our experimental setting is very low (less than 5). While it is not possible to formally determine if Shea tree experienced a simple expansion or underwent a bottleneck followed by an expansion, the expansion signal appear to be very strong as it involves a size change of 70.0 time on average considering scenario 3 and of 158.0 time on average considering scenario 4. We may however try to propose explanations to understand why these scenarios can't be distinguished in our data. On the first hand, considering additional analysis of pseudo observed data sets, scenario 3 and 4 are often confounded. This suggests that our experimental design is not powerful enough to distinguish between the 2 scenarios. On the other hand, we assumed in our simulations that Shea tree in West Africa was a single panmictic population and we did not account for any spatial component of the data – the expansion is assumed to be purely demographic. This is of course very simplistic as there is definitely a pattern of isolation by distance and range expansion can't be excluded. This may generate a more clustered genealogy (Excoffier *et al.*, 2009) that can blur the expansion signal or even be confounded with a bottleneck (Reich & Goldstein, 1998). In that context, the bottleneck signal detected in scenario 4 could be a pure artefact (Figure II.8). An additional burning question is the expansion time because it could give cues on important historical events having affected the West Africa savannah biome. Unhappily, the posteriors for expansion time are sufficiently broad not to allow determining whether the expansion was caused by Shea tree domestication (between 5000-1000 BP) (Lovett & Haq, 2000; Maranz & Wiesman, 2003) or by some major climatic events (last glacial maximum 120000 BP) (Maley, 1989; Maley, 1991; Maley, 1996).



Chapitre II : Étude des Processus historiques et démographiques



***Chapitre III :
Variabilité des
constituants
chimiques au sein
des graines du
Karité***



1 INTRODUCTION

1.1 Rôle adaptatif des réserves des graines

Durant la germination, avant l'initiation de la photosynthèse, les angiospermes et les gymnospermes utilisent exclusivement les réserves stockées dans leurs graines. La quantité et la qualité de ces réserves sont très importantes en raison de leur rôle dans la germination. Cette période (la germination), bien que courte dans la vie des plantes, joue pourtant un rôle déterminant dans la fitness, c'est-à-dire la capacité d'un individu à assurer sa descendance (Rees & Long, 1992; Westoby *et al.*, 1992). Chez les espèces pérennes en particulier, certaines émergent à l'ombre de la canopée, la quantité et la qualité des réserves de leurs graines ont donc un rôle important en attendant leur autonomie en énergie et en carbone permise par la photosynthèse (Westoby *et al.*, 1992). De plus, chez le Karité notamment, comme chez le chêne ou encore le maïs, la germination est dite « hypogée » (Hall *et al.*, 1996). Ce mode de germination implique que la graine reste sous la surface du sol, l'épicotyle s'allongeant alors plus vite que l'hypocotyle. Ce mode de germination associé à une croissance lente (chez le Karité notamment), implique une période longue durant laquelle la jeune plantule reste donc dépendante des réserves disponibles dans la graine avant d'atteindre son autonomie photosynthétique.

Ainsi, dans la mesure où la quantité et la qualité des réserves stockées dans la graine jouent un rôle déterminant lors de la germination et dans la fitness des plantes, ces caractéristiques sont fortement soumises à la sélection naturelle (Linder, 2000). Dans les réserves des graines nous nous intéressons dans cette thèse aux composés lipidiques et parmi eux, aux acides gras et aux tocophérols.

1.2 Les composés lipidiques des graines

Une importante mise en réserve de composés lipidiques est observée au sein des graines, notamment chez les plantes oléagineuses (c'est le cas du Karité), et au sein de tissus pérennants tels que les rhizomes ou les rayons médullaires. Ce stockage de lipides constitue la forme la plus condensée d'accumulation du carbone (Guignard *et al.*, 1985; Somerville *et al.*, 2000) nécessaire pour assurer la germination et la croissance post-germinative (Huang, 1996). Cependant, ces composés n'ont pas pour unique rôle l'apport en énergie et en



carbone. Certains d’entre eux ont un rôle de protection contre notamment les radicaux libres. En effet, les lipides présentent une grande variété moléculaire avec des propriétés différentes (Gunstone *et al.*, 1994; Somerville *et al.*, 2000). Si ces composés ne montrent pas d’unité structurelle, ils possèdent tous une propriété identique : ils sont insolubles dans l’eau (solubles dans le chloroforme et l’éther). Ces corps gras globalement hydrophobes se classent en deux grandes familles : les lipides saponifiables, et les lipides insaponifiables.

1.2.1 Les lipides saponifiables

Les lipides saponifiables constituent la quasi-totalité des lipides des graines. Ils se trouvent sous deux formes dans les graines : les acides gras libres ; et les triacylglycérols, également appelés triglycérides ou TAG (Gunstone *et al.*, 1994; Somerville *et al.*, 2000).

Tableau III.1. Principaux acides gras des graines du Karité.

Nom commun	Nom systématique	Structure	Abréviation
ACIDES GRAS SATURÉS			
Acide palmitique	Acide <i>n</i> -hexadécanoïque		16 : 0
Acide stéarique	Acide <i>n</i> -octadécanoïque		18 : 0
Acide arachidique	Acide <i>n</i> -éicosanoïque		20 : 0
ACIDES GRAS INSATURÉS			
Acide oléique	Acide <i>cis</i> -9-octadécénoïque		18 : 1 ^{Δ9}
Acide vaccénique	Acide <i>cis</i> -7-octadécénoïque		18 : 1 ^{Δ7}
Acide linoléique	Acide <i>cis,cis</i> -9,12-octadécatriénoïque		18 : 2 ^{Δ9,12}
Acide γ-linolénique	Acide <i>tout-cis</i> -6,9,12-octadécatriénoïque		18 : 3 ^{Δ6,9,12}

Les acides gras libres sont très minoritaires dans les graines, ils ne représenteraient chez le karité que 1 à 20% de la matière grasse (Okullo *et al.*, 2010). Ils constituent cependant le substrat indispensable à la formation des autres lipides saponifiables (**Figure III.1**). La synthèse de ces composés s’initie au sein du stroma plastidial des cellules végétales avec la mise en place d’un cycle d’élongation progressif par l’ajout de deux carbones d’un thioester : le malonyl-coenzyme A (Somerville *et al.*, 2000). Les acides gras ainsi formés chez les plantes possèdent 16 ou 18 carbones, mais une élongation secondaire de la chaîne carbonée de ces composés est possible après exportation vers le réticulum endoplasmique (Somerville *et al.*, 2000). Les sept principaux acides gras présents dans la graine de karité (Maranz & Wiesman, 2003; Maranz *et al.*, 2004) sont résumés dans le **Tableau III.1**. Ce sont

également les acides gras les plus couramment rencontrés dans toutes les plantes et tout type de tissu confondu (Somerville *et al.*, 2000). Les acides gras se distinguent en deux groupes majeurs qui sont les acides gras saturés, c'est-à-dire dont la chaîne carbonée est complètement hydrogénée, et les acides gras insaturés présentant une ou plusieurs doubles liaisons au sein de leur chaîne hydrocarbonée (**Tableau III.1**). Cependant, comme nous venons de le voir, les acides gras sont rarement à l'état libre dans les graines mais complexés au sein de molécules lipidiques saponifiables plus grosses telles que les triglycérides (Somerville *et al.*, 2000).

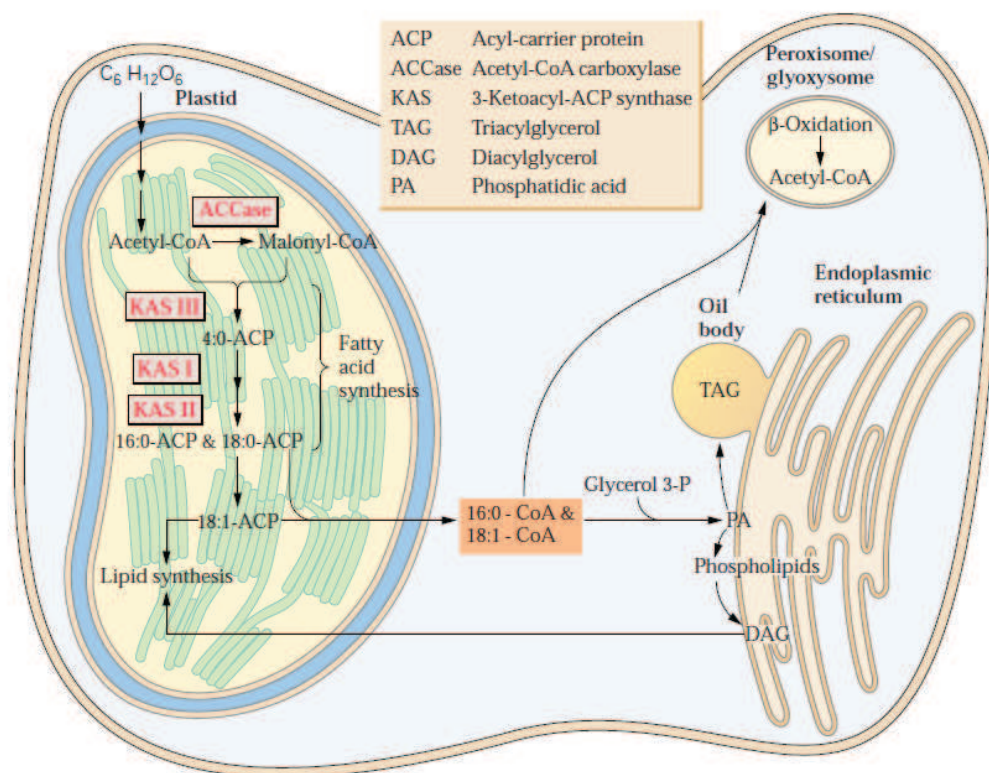


Figure III.1. Voie de biosynthèse simplifiée des lipides

Les triglycérides constituent les composants majeurs des lipides de la graine. Ces lipides saponifiables résultent de la triple réaction d'estérification entre un trialcool (le glycérol) et trois acides gras identiques ou différents (**Figure III.2**). La biosynthèse des TAG a lieu dans le réticulum endoplasmique, puis les triacylglycérols sont libérés dans le cytosol sous forme de gouttelettes lipidiques ou dans la lumière du réticulum endoplasmique (**Figure III.1**) (Somerville *et al.*, 2000). La structure moléculaire de ces composés leur donne un potentiel de libération d'énergie sous forme d'adénosine triphosphate (ATP) au moins deux fois supérieure à celle fournie par les composés glucidiques (Somerville *et al.*, 2000). Ils

apportent ainsi à la graine, l'énergie nécessaire pour la germination (Somerville *et al.*, 2000). Ces composés non polaires présentent une température de fusion qui s'élève avec le niveau de saturation en hydrogène et la longueur de la chaîne hydrocarbonée des acides gras qui les composent (Bailey, 1982). Il existe également d'autres formes de lipides saponifiables dans les plantes : les phospholipides, les galactolipides et les sulfolipides. Cependant, ils se trouvent principalement dans les membranes plasmiques et non dans les réserves des graines (Somerville *et al.*, 2000).

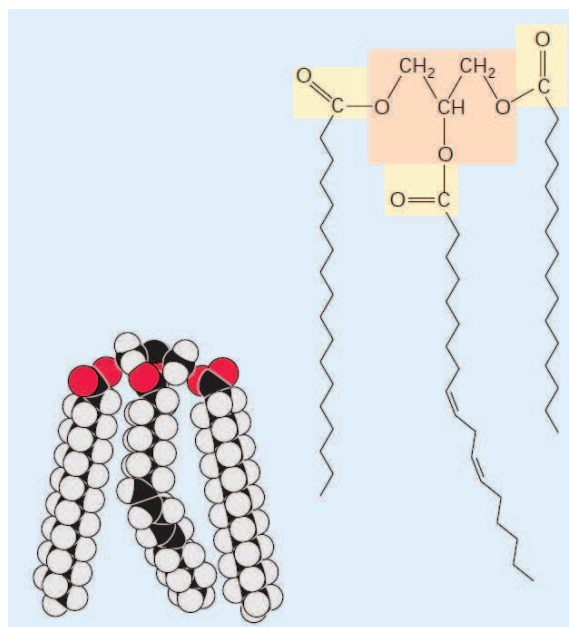


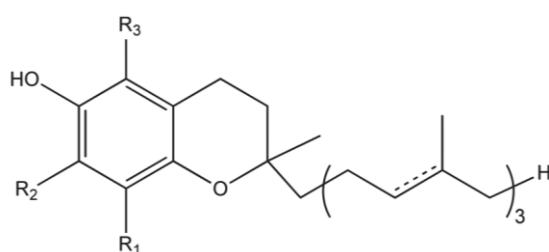
Figure III.2. Exemple d'un triglycéride : stéarique-linoléique-palmitique

1.2.2 Les lipides insaponifiables

Les lipides insaponifiables sont présents en très faibles quantités chez l'ensemble des végétaux et représentent moins de 1% des lipides totaux. Cependant, ces composés dérivés de l'isoprène (ou 2-méthylbuta-1,3-diène) jouent un rôle biologiquement fondamental (hormones, vitamines). Ils se divisent en deux sous-groupes qui sont les stéroïdes et les terpènes (Vollhardt & Schore, 2004). Nous nous concentrons pour cette thèse uniquement sur le groupe des composés terpéniques. Ils sont des polymères linéaires ou cycliques d'isoprènes. Selon le nombre d'entités isoprène incorporées dans leur structure, les terpènes sont subdivisés en monoterpènes (C₁₀ = 2 entités isoprène), sesquiterpènes (C₁₅ = 3 entités isoprène), diterpènes (C₂₀ = 4 entités isoprène), triterpènes (C₃₀ = 6 entités isoprène) et tetraterpènes (C₄₀ = 8 entités isoprène) (Vollhardt & Schore, 2004). Ces composés terpéniques chez les plantes regroupent notamment les composants aromatiques

des huiles essentielles telles que les santalols (Bottin *et al.*, 2007), les pigments comme les caroténoïdes (pigments jaune à rouge retrouvés par exemple dans les carottes) ou les lycopènes (pigments rouges des tomates) et les vitamines liposolubles avec notamment les vitamines E (tocophérols et tocotriénols) (Vollhardt & Schore, 2004).

Dans cette thèse nous nous intéressons en particulier aux tocophérols. Ces composés diterpéniques forment avec les tocotriénols, le groupe de tocochromanols, plus couramment appelé Vitamine E. Dans les tocophérols, contrairement aux tocotriénols, la chaîne latérale C16 est saturée, c'est-à-dire complètement hydrogénée, et le degré de méthylation du cycle benzénique détermine les quatre formes de tocophérols : alpha (5,7,8-triméthyl), bêta (5,8-diméthyl), gamma (7,8-diméthyl) et delta (8-méthyle) (**Figure III.3**).



	R1	R2	R3
α -tocophérol	CH ₃	CH ₃	CH ₃
β -tocophérol	CH ₃	H	CH ₃
γ -tocophérol	CH ₃	CH ₃	H
δ -tocophérol	H	H	H

Figure III.3. Structure chimique des tocophérols

Le mécanisme de biosynthèse des tocophérols est bien compris (**Figure III.4**) (Gunstone *et al.*, 1994). Il implique le couplage d'un phytyl diphosphate à l'acide homogentisique (acide 2,5-di-hydroxy-phénylacétique), suivie par des réactions de cyclisation et de méthylation (**Figure III.4**) (Gunstone *et al.*, 1994; Collakova & Dellapenna, 2003). Cette synthèse s'effectue dans le chloroplaste des plantes, où la tyrosine (un acide aminé aromatique) peut être considérée comme le précurseur de base de cette voie de biosynthèse (**Figure III.4**). À la première étape, cet acide aminé est oxydé en acide *p*-hydroxy-pyruvique, ensuite converti en acide homogentisique par l'enzyme *p*-hydroxy-phényl pyruvate dioxygénase. L'acide homogentisique est ensuite condensé avec le phytyl diphosphate dans une réaction catalysée par l'enzyme homogentisate phytyl transférase pour donner le 2-méthyl-6-phytyl-1,4-benzoquinol. La voie de biosynthèse se subdivise à ce niveau avec : soit une étape de cyclisation aboutissant à la synthèse des δ -tocophérols ensuite méthylés en β -tocophérols ; soit une étape de méthylation aboutissant au 2,3-diméthyl-5-phytyl-1,4-benzoquinol devenant le substrat à la synthèse du γ -tocophérol lui-même méthylé en α -tocophérol (**Figure III.4**).

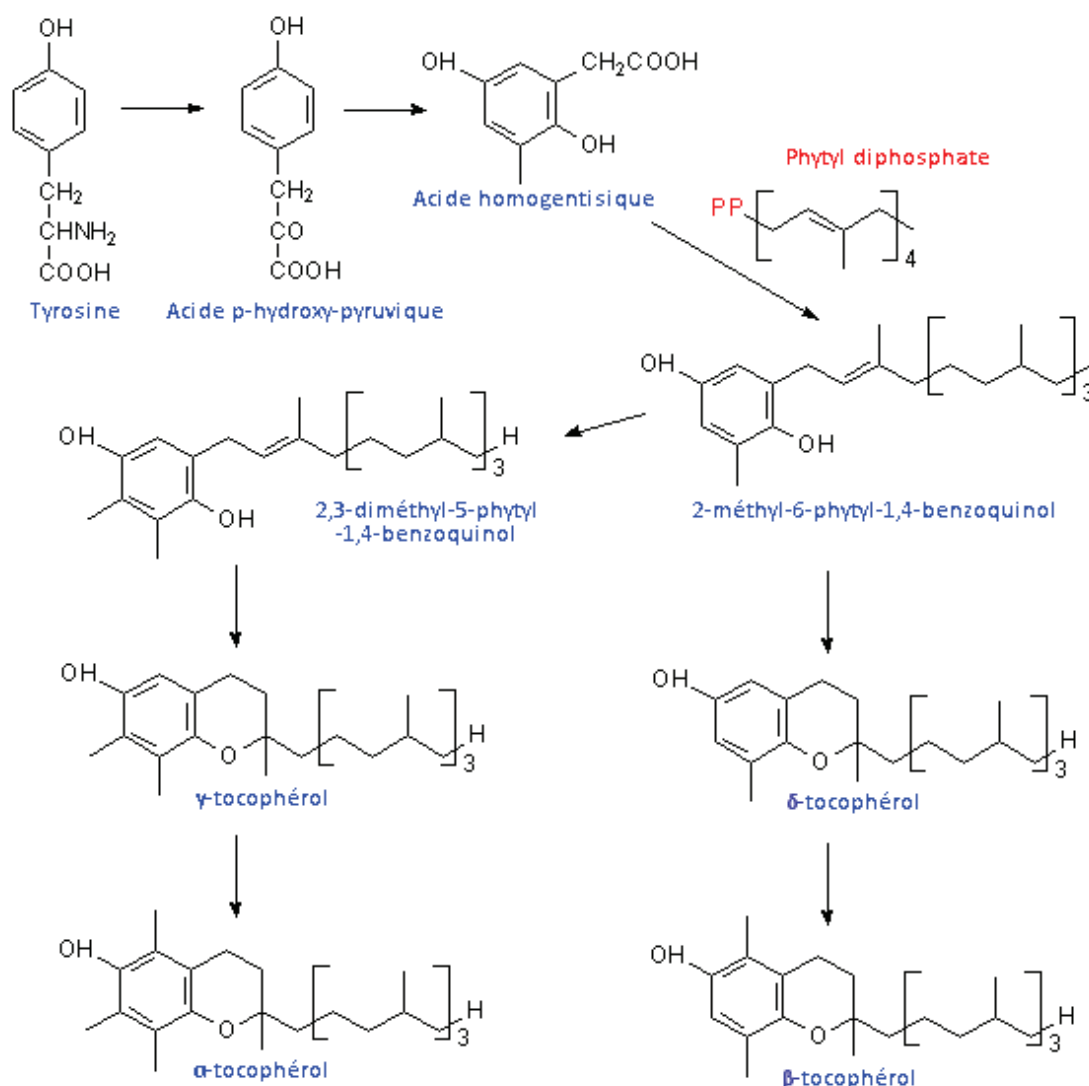


Figure III.4. Voie de biosynthèse des tocophérols

Le mécanisme d'action antioxydante de ces composés est de type « *chain-breaking* » (Kamal-Eldin & Appelqvist, 1996a; Brigelius-Flohé, 2009). Les tocophérols appartiennent à cette catégorie d'antioxydants en tant que « donneurs d'hydrogène » et permettent d'enrayer les réactions en chaînes de peroxydation des lipides notamment. Ainsi, après une réaction d'oxydation par exemple d'un acide gras $\text{CH}_3(\text{CH}_2)_n\text{-COOH} + \text{OH}^\bullet \rightarrow \text{CH}_3(\text{CH}_2)_n\text{-COO}^\bullet + \text{H}^+$, le tocophérol va donner un hydrogène et être transformé en radical tocophéroxyyle stable afin de bloquer la propagation radicalaire. Ce radical évolue ensuite vers un composé d'oxydation non radicalaire tel que la tocophérylquinone (Kamal-Eldin & Appelqvist, 1996a; Brigelius-Flohé, 2009). Si ce mécanisme de « *chain-breaking* » consomme les tocophérols, un système de recyclage de ces composés, par l'intermédiaire de l'acide ascorbique notamment, leur permettent de retrouver leur potentiel antioxydant (Kamal-



Eldin & Appelqvist, 1996a; Brigelius-Flohé, 2009). Parmi les quatre formes de tocophérols, l' α -tocophérol, étant le plus méthylé, possède le pouvoir antioxydant le plus puissant (Brigelius-Flohé, 2009).

1.3 Déterminisme génétique et environnemental de la variabilité des constituants chimiques des graines

La variabilité des constituants chimiques au sein des graines résulte d'effets génétiques et environnementaux. La part de ces effets dépend fortement des espèces et des constituants et a surtout été étudiée pour quelques espèces de grande culture telles que le soja (Rebetzke *et al.*, 1998; Primomo *et al.*, 2002; Hou *et al.*, 2006); le colza (Goffman & Mollers, 2000; Goffman & Becker, 2001b, a; Marwede *et al.*, 2004), ou le tournesol (Velasco *et al.*, 2002), ou pour des espèces modèles telles qu'*Arabidopsis thaliana* (Gilliland *et al.*, 2006). Les principales études montrent que l'héritabilité de certains acides gras varie de 15 à 88% (Schierholt & Becker, 2001; Hou *et al.*, 2006; Bachlava *et al.*, 2007; Aruna & Nigam, 2009). De même l'héritabilité du taux des tocophérols varie de 24 à 44% (Goffman & Becker, 2001b; Velasco *et al.*, 2002; Marwede *et al.*, 2004; Amaral *et al.*, 2005; Gilliland *et al.*, 2006). La valeur de ces héritabilités montre que ces caractères (nous faisons allusion aux acides gras et aux tocophérols) sont relativement influencés par l'environnement mais que leur variabilité relève du polymorphisme de gènes impliqués dans l'expression de ces caractères. Cependant, l'héritabilité modérée observée montre que ces caractères sont aussi gouvernés par un nombre de gènes importants (Gallais, 1990).

La variation d'origine génétique dépend fortement de la population concernée et donc de la fréquence allélique, et celle estimée à partir d'un échantillon restreint à une petite population est attendue moins élevée que celle issue d'un échantillon couvrant toute l'aire. Pour les espèces forestières, qui couvrent des aires très large, il apparait nécessaire de comprendre la façon dont s'organise la variabilité selon différents niveaux hiérarchiques : continents, régions, sous régions, méta-populations, populations. L'étude d'autres types de caractères sous contrôle polygéniques liés notamment à la croissance ou à la production de biomasse, ont montré qu'une part importante de la variabilité au sein des espèces pouvait se situer entre des grandes zones de distribution de l'aire, conséquences de plusieurs



Chapitre III : Variabilité des constituants chimiques des graines

facteurs évolutifs (dérive, sélection naturelle) favorisant cette différenciation (McKay & Latta, 2002; Le Corre & Kremer, 2003; Latta, 2004).

En complément de la compréhension de la part des effets environnementaux et génétiques dans la variation au sein des graines, notre objectif est aussi de comprendre comment se distribue cette variabilité au sein de l'aire naturelle pour les constituants chimiques des graines.

1.4 Variation en relation avec les gradients environnementaux

L'origine de cette variation génétique reste mal connue. Est-elle le résultat d'adaptation locale aux conditions environnementales (facteurs abiotiques tels que le climat, les conditions édaphiques) ? Est-elle le résultat d'adaptation aux conditions écologiques et notamment aux facteurs biotiques tels que : les disséminateurs des graines, la pression des pathogènes et des prédateurs ?

Chez les espèces pérennes, certaines études portant sur des caractères de croissance liés à l'adaptation ont montré pour des essences forestières, une relation forte entre les conditions environnementales de l'origine des populations et la variation d'origine génétique, selon un processus clinal. C'est le cas pour l'*Eucalyptus urophylla* distribué sur un gradient altitudinal dans les îles de la Sonde au Timor (Triapiana *et al.*, 2007). Dans le cas des constituants chimiques des graines quels patrons de variation observe-t-on sur des gradients environnementaux marqués (température, pluviométrie, etc...) ?

Peu d'études se sont intéressées aux bases de la variation des constituants chimiques des graines au sein des espèces végétales. Une étude menée au sein du genre *Helianthus* (Linder, 2000) a montré notamment que les espèces évoluant à des latitudes chaudes présentent des huiles plus riches en acides gras saturés, tandis que les espèces des latitudes plus tempérées montrent des profils plus riches en acides gras insaturés. Ce constat s'explique par le fait que des proportions d'acides gras insaturés plus importantes (bien que libérant moins d'énergie) permettent dans un milieu plus frais de germer plus rapidement, pousser plus vite et donner un avantage sélectif aux plantules dans des milieux très concurrencés. À l'inverse, les plantules présentant des proportions plus importantes en



acides gras saturés possèdent des réserves d'énergies plus importantes leur conférant un avantage sélectif afin de s'établir sur des sols moins favorables (Linder, 2000).

1.5 Patron de variabilité chez le Karité

Dans le cas du karité l'analyse de la variabilité des constituants chimiques des graines a été abordée au cours du projet « ProKarité » (<http://www.prokarite.org/>), à partir d'échantillons récoltés en milieu naturel sans expérimentation permettant de séparer les effets environnementaux des effets génétiques. Ce projet ProKarité (projet d'appui technique à la filière Karité) a pour objectif l'optimisation de la gestion de la production et le contrôle qualité en vue de la certification internationale des produits issus de cette filière. Les études publiées à la suite de ce projet ont donc consisté essentiellement à décrire les grands patrons de variations des constituants chimiques majeurs de la graine de karité sur son aire naturelle avec l'élaboration d'une première base de données chimiques « *Vitellaria* » (<http://www.prokarite.org/vitellaria-dbase/index.html>) pour les acides gras (Maranz & Wiesman, 2003; Maranz *et al.*, 2004; Di Vincenzo *et al.*, 2005), les composées phénoliques (Maranz *et al.*, 2003) et les tocophérols (Maranz & Wiesman, 2004).

Les conclusions apportées par ces études ont permis de confirmer des résultats connus tels que la différence des teneurs en acide oléique et stéarique entre les populations de karité d'Afrique de l'Est et d'Afrique de l'Ouest (Hall *et al.*, 1996). Elles ont fourni des résultats originaux concernant la relation entre la teneur en acide gras et les gradients climatiques entre la partie sahélienne du nord de l'aire du karité (faible pluviométrie et forte température) et la partie soudano-guinéenne du sud de l'aire (forte pluviométrie température plus modérée) (Maranz & Wiesman, 2003). De même il a été montré une forte variation Nord-Sud pour le taux de tocophérol, multiplié par 15 entre les deux extrémités de l'aire (Maranz & Wiesman, 2004).

Ces analyses pionnières laissent penser que la variation environnementale peut avoir une forte influence sur la teneur en constituants chimiques des graines et pourrait agir comme une pression de sélection ayant favorisé certains allèles et donc créé un déterminisme génétique marqué de ces constituants. Ceci étant, les travaux publiés à partir de la base de données « ProKarité » (Maranz & Wiesman, 2003; Maranz *et al.*, 2003; Maranz & Wiesman, 2004; Maranz *et al.*, 2004; Di Vincenzo *et al.*, 2005) ont été réalisés sur un



Chapitre III : Variabilité des constituants chimiques des graines

échantillonnage réduit et surtout très déséquilibré. Par exemple, pour l'étude portant sur l'influence de la température sur la variation des teneurs en tocophérols (Maranz & Wiesman, 2004), il apparaît que 11 sites sur les 21 concernés par l'étude, ne présente que 1 à 2 individus (**Figure III.12** et **III.13**). De plus, le nombre de graines analysées par arbre semble parfois très faible (1 à 3 graines).

Afin de répondre en partie aux questions posées ci-dessus et pallier certaines faiblesses des dispositifs d'échantillonnage précédents, nous avons développé une étude à partir d'un échantillon couvrant à la fois les parties les plus distantes de l'aire naturelle du karité, mais aussi suivant un gradient climatique nord-sud. Par ailleurs nous avons établi un échantillonnage impliquant un nombre de fruits par arbre (entre 20 et 30) et un nombre d'arbres par populations (entre 12 et 40) beaucoup plus important que dans les études précédentes. Cependant, de la même manière que pour les études précédentes, notre échantillon étant réalisé à partir de populations naturelles, il ne sera pas possible d'estimer des héritabilités et donc de réellement mesurer la part des effets génétiques par rapport aux effets environnementaux.

L'objectif de ce chapitre est de comprendre les patrons de variation de certains constituants chimiques des graines (acides gras et tocophérols) au sein de l'aire naturelle et leur relation avec certains paramètres environnementaux.



2 MATÉRIEL & MÉTHODES

2.1 Sites échantillonnés, mesures morphologiques et caractéristiques environnementales

Afin d'étudier la variabilité chimique et morphologique des propagules du Karité un échantillonnage sur 17 sites dans 5 pays de l'aire naturelle (Sénégal, Mali, Ghana, Burkina Faso et Ouganda) a été effectué en suivant les protocoles établis pour le projet européen INCO INNOVKAR (Innovative Tools and Techniques for Sustainable Use of the Shea Tree in Sudano-Sahelian zone) (**Figure III.5**). Ainsi, au sein de chaque site de récolte une distance de 50 à 100 mètres a été respectée entre les arbres. Pour chaque arbre un code a été défini, l'altitude et le positionnement GPS ont été répertoriés et les caractéristiques suivantes ont été mesurées : diamètre à hauteur de poitrine (DBH = *Diameter at Breast Height*), hauteur de l'arbre (Hteur) et diamètre du houppier (DMH). Les arbres présentant des DBH inférieurs à 15 cm n'ont pas été choisis afin de ne pas sélectionner des arbres trop jeunes dans l'échantillonnage. Pour chaque arbre, 20 à 40 fruits mûrs ont été récoltés. Les fruits ont ensuite été déulpés et les noix séchées durant trois jours à 60°C en étuve dans les zones de récolte. À l'arrivée au laboratoire du CIRAD à Montpellier les noix ont été séchées à nouveau afin de stabiliser leur teneur en eau, puis stockées à température ambiante avant analyse. Pour chaque fruit de chaque arbre échantillonné les mesures morphologiques suivantes ont été effectuées : longueur du fruit (LongFruit), largeur du fruit (LargFruit), poids total du fruit (PTF), poids de la noix fraîche (PNF) et poids de la noix sèche (PNS). Cependant, ces mesures n'ont pu être pleinement collectées que pour les sites du Mali (Nafégué, Mperesso, Daelan, Tori et Sassambourou) du fait des difficultés sur site d'avoir l'appareillage et/ou les moyens humains nécessaires.

Pour certains sites, deux prélèvements ont pu être effectués sur deux années consécutives 2007 et 2008 : Nafégué, Daelan et Mperesso au Mali ; Pader et Katakwi en Ouganda. Cependant de nombreux arbres récoltés en 2007 n'ont pas fructifiés en 2008 et d'autres arbres ont donc été récoltés (en particulier en Ouganda et à Nafégué au Mali). Au



Chapitre III : Variabilité des constituants chimiques des graines

total sur les deux années de récoltes 2007 et 2008, 624 prélèvements ont été effectués et plus de 5000 fruits ont été caractérisés sur le plan morphologique.

La carte (**Figure III.5**) montre la distribution des sites échantillonnés sur l'aire naturelle du Karité. Tous les sites sélectionnés se trouvent au sein de systèmes agroforestiers (**Figure III.6**) organisés en cycle champ/jachère permettant une régénération naturelle durant la période de friche que constitue la jachère (Kelly *et al.*, 2007). Cet équilibre est décrit plus avant dans le Chapitre I « *le Karité dans l'écosystème de savane africaine* ».

Les **Tableaux III.3** et **III.4** résument les caractéristiques environnementales des sites. En plus de l'altitude et la distribution longitudinale et latitudinale des populations, 20 variables climatiques ont été utilisées pour caractériser les sites de récolte en utilisant la base de données WorldClim (Hijmans *et al.*, 2005) incluant : le nombre de mois secs (DRY_M) défini par le nombre de mois dont des précipitations (en mm) ne dépassent pas 2 fois la température moyenne mensuelle (en °C); la température annuelle moyenne (ANN_MEAN_T) ; l'écart de température mensuel moyen (TEMP_RANGE) ; la saisonnalité des températures (TEMP_SEAS) définie par l'écart-type de température; la température maximale du mois le plus chaud (MAX_T) ; la température minimale du mois le plus froid (MIN_T) ; l'écart de température annuel moyen (ANN_T_RANGE) défini par T_MAX moins T_MIN; l'isothermalité (ISOTHERM) définie par le rapport entre TEMP_RANGE et ANN_T_RANGE; la température moyenne du trimestre le plus humide (MEAN_T_WtQ) ; la température moyenne du trimestre le plus sec (MEAN_T_DQ) ; la température moyenne du trimestre le plus chaud (MEAN_T_WaQ) ; la température moyenne du trimestre le plus froid (MEAN_T_CQ) ; les précipitations annuelles (ANN_PRC) ; les précipitations du mois le plus humide (PRC_WtM) ; les précipitations du mois le plus sec (PRC_DM) ; la saisonnalité des précipitations (PRC_SEAS) définie par le coefficient de variation des précipitations annuelles ; les précipitations du trimestre le plus humide (PRC_WtQ) ; les précipitations du trimestre le plus sec (PRC_DQ) ; les précipitations du trimestre le plus chaud (PRC_WaQ) ; et les précipitations du trimestre le plus froid (PRC_CQ). On observe que pour la majeure partie de ces variables climatiques, les variations observées s'expriment entre l'Afrique de l'Ouest (sites 1 à 13) et l'Afrique de l'Est (sites 14 à 17).



Chapitre III : Variabilité des constituants chimiques des graines

Afin de décrire au mieux les caractéristiques édaphiques des sites, les types de sols ont été récupérés sur la base de données FAO-SIG (Oldeman & Van Engelen, 1993). Cinq types de sols sont représentés dans notre échantillonnage (**Tableau III.2**) :

- *LEPTOSOLS* : ces sols sont caractéristiques de paysages en érosion, très pauvres, de faible profondeur ou à teneur très élevée en graviers. (Indice de Fertilité : 1)

- *FERRALSOLS* : ces sols sont extrêmement altérés. D'origine du Pléistocène, ils sont caractérisés par la dominance d'argile kaolinique, une accumulation résiduelle de fer d'oxydes et d'hydroxydes d'aluminium, une structure de sol stable, un ratio limon/argile faible et une très faible teneur en minéraux altérables. Ils sont profonds mais pauvres chimiquement, avec une capacité d'échange d'ions faible, et des réserves de nutriments qui sont facilement perturbés par les pratiques agricoles. Ces sols nécessitent des cultures pérennes. (Indice de Fertilité : 2)

- *ACRISOLS* : les Acrisols sont des sols argileux de type principalement kaoliniques. Ils sont fortement altérables et sont souvent victimes de toxicité à l'aluminium après déforestation. Le potentiel limité de ces sols s'accommode bien aux agricultures itinérantes agroforestières basées sur des cycles de régénération champs/jachères de plusieurs dizaines d'années. (Indice de Fertilité : 3)

- *LIXISOLS* : les Lixisols sont des sols ferrugineux aux caractéristiques similaires aux Acrisols mais plus riches en nutriments et dépourvus de toxicité en Aluminium. Ils sont cependant également sujets à de fortes érosions et la FAO recommande des cultures pérennes pour ces sols. (Indice de Fertilité : 4)

- *VERTISOLS* : ces sols sont riches formés de sédiments ou roches altérées et d'argile de type « smectite », qui se gonfle lorsqu'il est mouillé. Ces sols sont souvent trouvés à des altitudes élevées ou dans des topographies ondulées.

Tableau III.2. Description des sites échantillonnés

Région	Code site	Nom site	GPS	altitude (m)	N* 2007 2008	Sous- espèce	Type de Sol	Type de parcelle	Hteur Moy	DBH Moy
AFRIQUE DE L'OUEST	SÉNÉGAL									
	1	Kenito	12.57N; 12.16O	161	12	<i>paradoxa</i>	<i>LEPTOSOLS</i>	Champ/Jachère		
	2	Samecoutha	12.60N; 12.13O	126	18	<i>paradoxa</i>	<i>LEPTOSOLS</i>	Champ/Jachère		
	3	Saraya	12.83N; 11.75O	180	10	<i>paradoxa</i>	<i>LEPTOSOLS</i>	Champ/Jachère		
	MALI									
	4	Nafégue	10.51N; 5.98O	344	35	<i>paradoxa</i>	<i>ACRISOLS</i>	Champ/Jachère		
	5	Mperesso	12.28N; 5.33O	340	34	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
	6	Daelan	13.25N; 4.99O	282	34	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
	7	Tori	13.61N; 3.72O	377	35	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
	8	Sassambourou	14.31N; 3.51O	392	35	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
	BURKINA FASO									
	9	Titao	13.72N; 2.16O	336	18	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
	10	Guibare	13.07N; 1.61O	303	21	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
	GHANA									
	11	Kulbia	10.83N; 0.96O	206	34	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
	12	Tolon	9.43N; 1.00O	154	35	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
	13	Kawampe	8.43N; 1.56O	125	35	<i>paradoxa</i>	<i>LIXISOLS</i>	Champ/Jachère		
AFRIQUE DE L'OUEST	OUGANDA									
	14	Katakwi	1.82N; 33.99E	1100	34	<i>nilotica</i>	<i>FERRALSOLS</i>	Champ/Jachère		
	15	Pader	2.80N; 33.31E	1031	31	<i>nilotica</i>	<i>FERRALSOLS</i>	Champ/Jachère		
	16	Moyo	3.62N; 31.64E	863	16	<i>nilotica</i>	<i>VERTISOLS</i>	Champ/Jachère		
	17	Uleppi-Arua	3.02N; 30.90E	1200	19	<i>nilotica</i>	<i>LEPTOSOLS</i>	Champ/Jachère		
Total				168	456					

* N : nombre d'arbres collectés

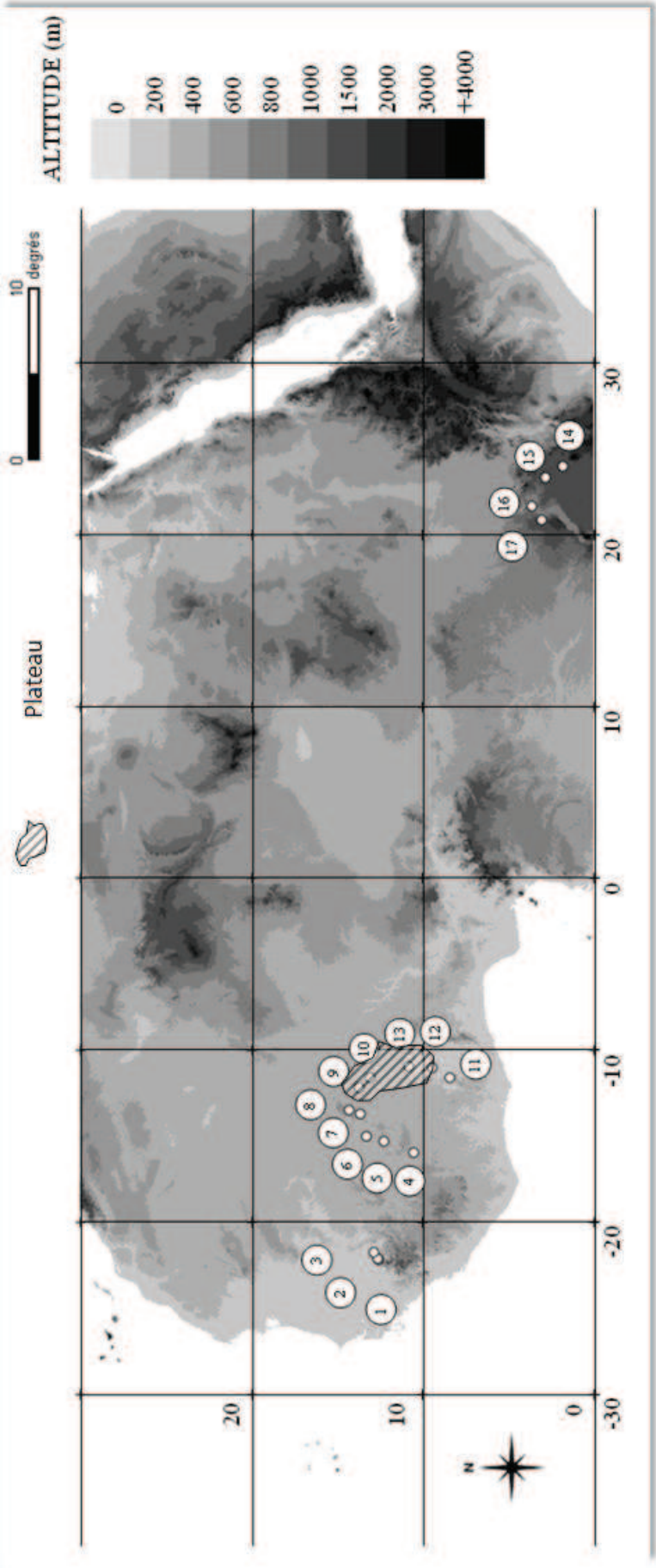
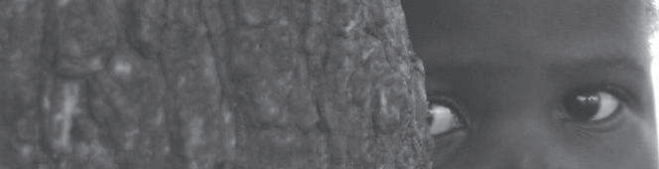
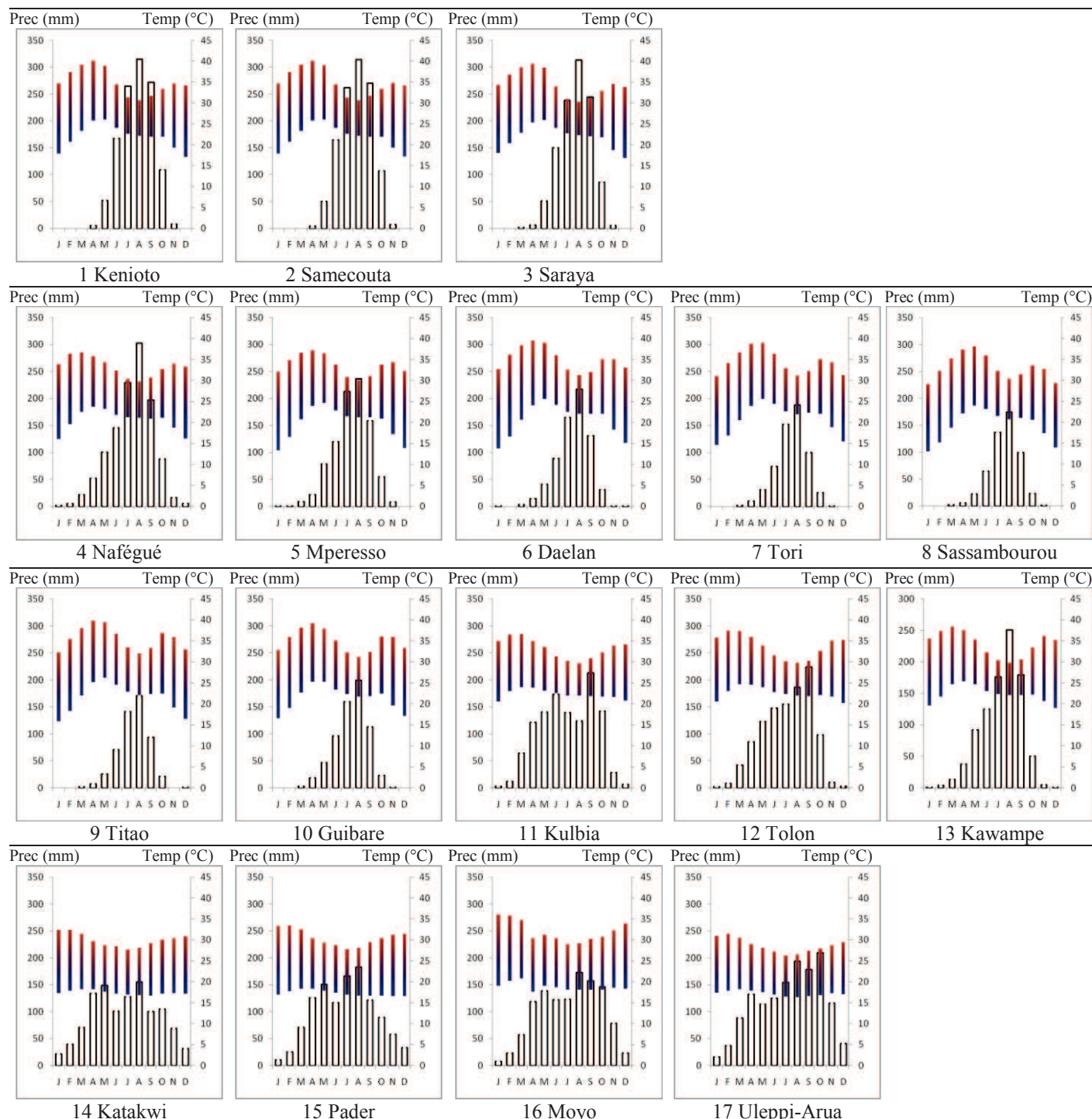


Figure III.5. Distribution des sites échantillonnés

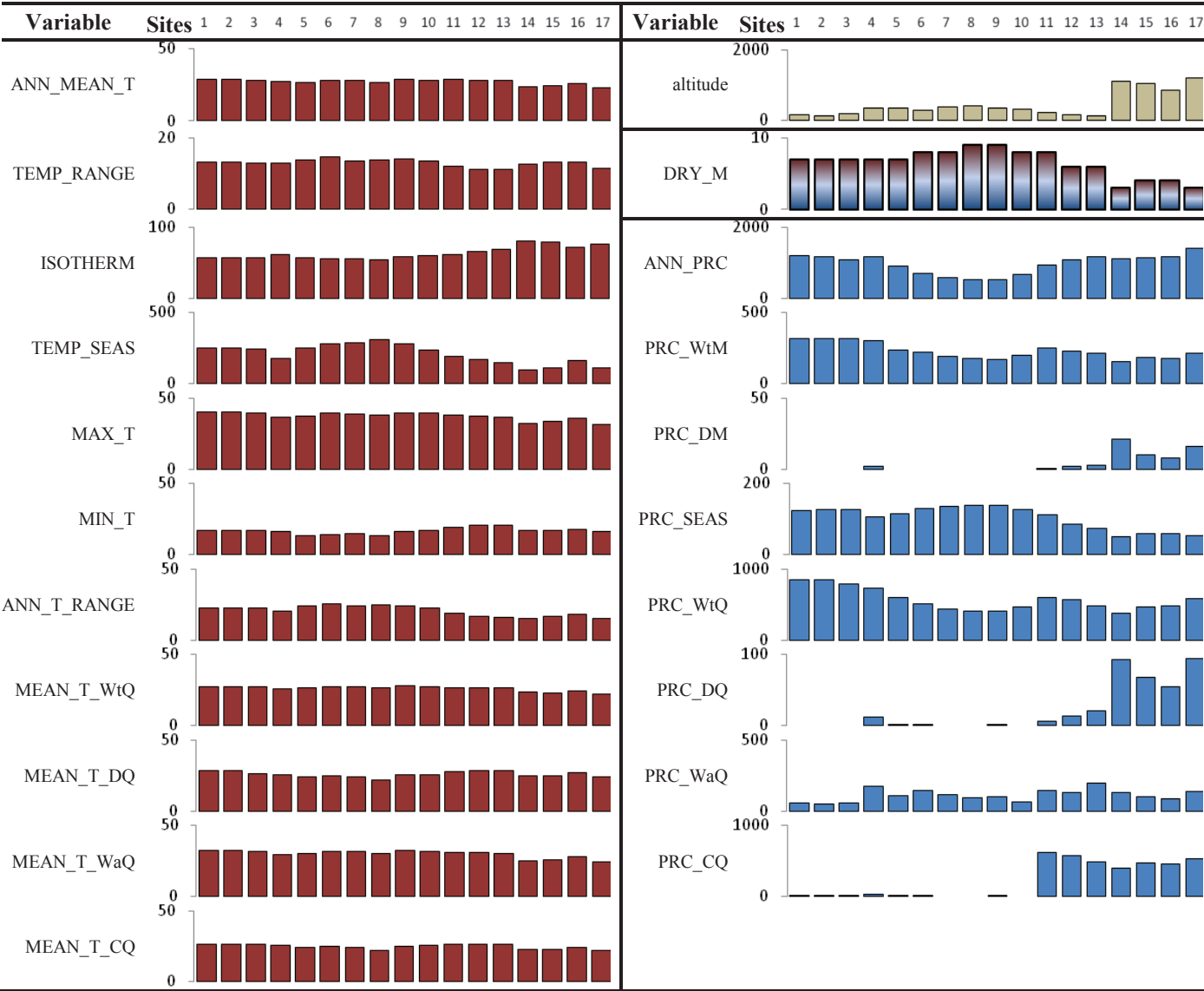
Tableau III.3. Températures moyennes et pluviométries mensuelles des sites échantillonnés





Chapitre III : Variabilité des constituants
chimiques des graines

Tableau III.4. Variation des sites selon diverses variables. La couleur verte correspond à l'altitude ; la couleur rouge correspond aux variables associées à la température ; la couleur bleu correspond aux variables associées à la pluviométrie.



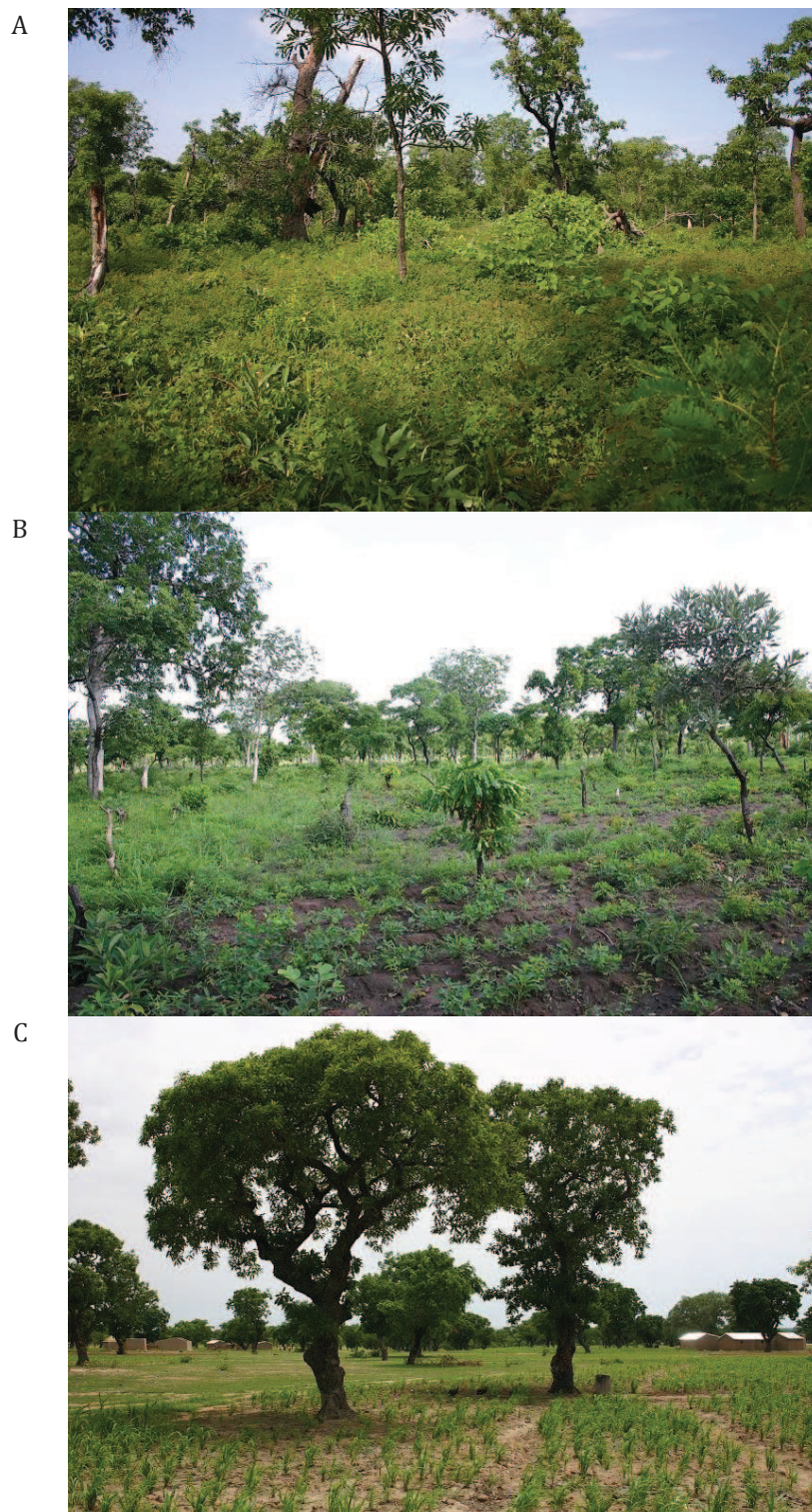
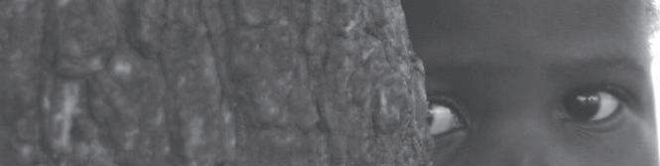


Figure III.6. Photographie de sites échantillonnés au Ghana : A = jachère près de Kawampe ; B = champ récemment exploité près de Kawampe ; et C = champ exploité depuis longtemps près de Kulbia.



2.2 Méthodes Analyse chimique

2.2.1 Préparation des échantillons

Une fois les noix de karité écosées, environ 15 amandes par arbre ont été broyées grossièrement en utilisant un robot “Vorwerk-Thermomix”. Ces poudres ont ensuite été congelées individuellement à -20°C puis broyées finement à l’aide d’un robot “SEB Valentin” afin d’obtenir une granulométrie finale comprise entre 0.5 et 0.8 mm. Les échantillons en poudre résultant ont ensuite été conservés à -20°C à l’abri de la lumière.

2.2.2 Erreur Standard de Laboratoire (ESL)

Pour chaque type d’analyse de laboratoire nous avons estimé l’erreur standard de laboratoire (ESL) en calculant l’écart type obtenu sur 10 répétitions d’analyses d’un échantillon standard de karité.

2.2.3 Mesure des teneurs en eau

La quantité d’eau contenue par les échantillons est mesurée par pesées différentielles (gravimétrie) après une incubation de 16 h dans une étuve (Chopin) à 103°C. Les teneurs en eau sont ensuite exprimées en pourcentage.

2.2.4 Extraction et mesure des teneurs en Matière Grasse (MG)

La matière grasse (MG) est extraite à l’éther de pétrole (solvant) pour chaque échantillon en utilisant un semi-automatique Soxtec™ 2050 extractor (FOSS analytique, A/S Slangerupgade 69, DK-3400 Hillerød, Danemark). L’huile contenue dans 4 g de poudre est ainsi extraite à 130 °C en suivant les étapes suivantes : extraction (1 h); rinçage (30 mn); évaporation du solvant (25 mn); séchage des cartouches d’extraction (5mn). Afin d’éliminer les dernières traces de solvant, les huiles extraites sont placées 30 mn en étuve à 60 °C. Le contenu de l’huile est déterminé par gravimétrie, puis exprimé en pourcentage de matière sèche ($MS = \text{masse de poudre initiale} - \text{masse d'eau}$). Les huiles ont enfin été conservées à -20 °C.

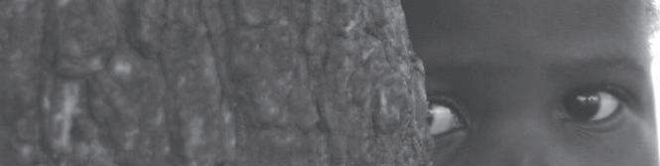
Cette méthode permet une extraction des huiles aussi efficace qu’avec les systèmes d’extracteurs classiques types « Soxlhet », tout en réduisant fortement la quantité de solvant



et en divisant par 5 le temps d'extraction (Schäfer, 1998). Cependant, afin d'en vérifier l'efficacité, nous avons comparé les résultats obtenus pour dix échantillons témoins extraits par cette méthode d'extraction « rapide » et par la méthode d'extraction conventionnelle. Aucune différence significative n'a pu être détectée entre les deux méthodes, que ce soit pour la quantité ou la caractérisation de la MG extraite.

2.2.5 Préparation des esters méthyliques d'acides gras (EMAG) et analyse des profils d'acides gras (AG)

Afin de caractériser les profils de composition des échantillons en acides gras (AG) par chromatographie en phase gazeuse (CPG), une étape préliminaire de transformation de la matière grasse en esters méthyliques d'acides gras (EMAG) a été effectuée. Pour cela, dans un ballon à fond rond de 25 ml, 10 mg d'huile sont ajoutés à 3 ml de solution de méthylate de sodium contenant une pointe de phénolphthaléine (un indicateur coloré acide-base). Le milieu réactionnel est porté au reflux pendant 10 min dans une première étape de méthanolyse basique durant laquelle les parties neutres glycériques des triglycérides sont transformées en esters méthyliques et les acides gras libres sont transformés en savons. Lors d'une seconde étape d'estérification acide, l'ajout de 3 ml de méthanol chlorhydrique induit une acidification du milieu conduisant à la décoloration de la phénolphthaléine, le mélange est à nouveau porté à reflux pendant 10 min provoquant une décomposition des savons précédemment formés en esters méthyliques. Le mélange est ensuite refroidi à température ambiante, puis 8 ml d'hexane et 10 ml d'eau sont ajoutés. La phase organique (hexane) contenant les EMAG formés, est récupérée pour analyse ultérieure en CPG. Les profils d'EMAG, représentatifs des compositions en AG des échantillons, sont déterminés par chromatographie en phase gazeuse (CPG) avec un appareil « Thermo Focus » (Thermo Fisher Scientific, 81, rue Wyman, Waltham, MA 02454). La colonne capillaire utilisée « CP SIL 88 » (Varian, Inc, 3120 Hansen Way, Palo Alto, CA 94304-1030) est très polaire et fortement substituée en cyanopropyl avec les caractéristiques suivantes : 30 m de longueur; 0,32 mm de diamètre intérieur; et 0,25 μm d'épaisseur du film. Le gaz vecteur utilisé est l'Hélium avec un débit de 1 ml/min et un ratio de division de 1/80. Les températures de l'injecteur-diviseur et du détecteur à ionisation de flamme sont respectivement 250 °C et 270 °C. La programmation de température du four s'établit comme suit : 1 min à 150 °C, de 150 °C à



225 °C à 5 °C/min, 2 min à 225 °C. Le volume d'EMAG injecté est de 1 µL. Une fois les chromatogrammes générés, pour chaque échantillon les pourcentages des surfaces de pics sont divisés par le poids moléculaire des EMAGs afin de déterminer le pourcentage molaire de chaque acide gras. Deux analyses sont exécutées par échantillon.

2.2.6 Mesure des teneurs en tocophérols

Pour chaque échantillon une extraction « solide-liquide » est effectuée à température ambiante. Dans un bécher de 25 ml, 3 g de poudre sont ajoutés à 12 ml l'éther de pétrole, puis mis sous agitation magnétique durant 2 h à l'obscurité, dans le but d'éviter la dégradation des composés thermolabiles tels que les tocophérols. Le solvant contenant la matière grasse est ensuite filtré par seringue PTFE 0,2 µm, puis évaporé à l'hélium sous hotte aspirante. Ces extraits de matière grasse sont conservés à -20 °C et à l'abri de la lumière. Les tocophérols contenus dans la matière grasse sont analysés par chromatographie en phase liquide à haute performance (CLHP) conformément à la norme ISO 9936 (Afnor, 1997). L'appareillage CLHP est constitué de modules fournis par Thermo-Finnigan (France) : une pompe quaternaire (P1000XR), un passeur d'échantillon (AS1000) et une valve d'injection à 6 voies muni d'une boucle de 20 µl, un détecteur spectrofluorimétrique (FL3000) et un logiciel pour le traitement des données (PC 1000). Les échantillons standards ont été achetés auprès de VWR International SAS France. La phase mobile consiste en un mélange hexane/dioxane (97 :3, v/v), avec un débit de 1 ml/min. La colonne de silice est une « Hypersil™ » de 5 µm d'épaisseur, 4 mm de diamètre interne par 250 mm de long. La longueur d'onde d'excitation est fixée à 290 nm et celle d'émission à 330 nm. Trois analyses sont exécutées par échantillon.

2.3 Analyses Statistiques

Tout comme les analyses chimiques, les traitements statistiques subséquents ont été réalisés en dissociant les caractères en deux groupes majeurs : profils en acides gras ; et teneur en tocophérols.



2.3.1 Analyses descriptives

Afin de décrire la variabilité chimique de la composition chimique des graines de Karité, une analyse statistique descriptive a été effectuée à différentes échelles : aire naturelle ; sous-espèces ; transects ; populations. Ainsi la moyenne, l'intervalle de variation et l'écart-type ont été calculés pour chaque constituant chimique en utilisant le logiciel XLStat 2009 (Addinsoft, Paris, France).

2.3.2 Analyses de corrélations

Les analyses de corrélation ont été effectuées grâce à la procédure CORR (SAS Institute 1989) sur les valeurs propres des individus.

2.3.3 Analyses des différences entre populations (sites)

Pour tous les constituants chimiques des graines du karité et certains traits morphologiques, nous avons testé la différence entre les groupes d'individus à différentes échelles : entre les régions ; entre les sites ; entre les sites au sein des régions et entre les sites au sein des transects. Dans ce but, nous avons testé l'hypothèse nulle H_0 d'égalité des moyennes entre les groupes par une analyse de variance en modèle fixe. Cette analyse a été réalisée grâce à la procédure GLM du logiciel statistique SAS (SAS Institute 1989) en précisant l'option « somme des carrées de type III », vérifiant l'hypothèse nulle H_0 par un test de Fisher. La comparaison des moyennes des groupes a été réalisée par un test de Bonferroni, le modèle suivant a été utilisé :

$$y_{ij} = \mu + \beta_j + e_{i(j)} \quad (1)$$

Avec :

y_{ij} , la valeur de la variable à expliquer pour l'individu i dans le groupe j ;

μ , la moyenne globale de la variable à expliquer ;

β_j , l'effet fixe du j -ième groupe ;

$e_{i(j)}$, l'erreur résiduelle d'effet aléatoire suivant la loi $\sim \mathcal{N}(0, \sigma_e^2)$ liée à l'individu i dans le groupe j



Analyse de variance hiérarchique en modèle aléatoire :

Afin de comprendre la structure de la variance chimique des graines de karité sur l'aire naturelle, nous avons analysé les composantes de cette variance pour chaque caractère par une ANOVA hiérarchique en modèle aléatoire. Cette analyse a été réalisée grâce à la procédure VARCOMP du logiciel statistique SAS (SAS Institute 1989) qui utilise le maximum de vraisemblance restreint (REML) et des sommes des carrés des écarts de type III pour tester l'hypothèse nulle H_0 d'égalité des variances relatives aux groupes et aux sous-groupes au sein des groupes grâce à un test de Fisher. Le modèle utilisé est le suivant :

$$y_{ijk} = \mu + \alpha_k + \beta(\alpha)_{j(k)} + e_{i(jk)} \quad (2)$$

Avec :

y_{ijk} , la valeur de la variable à expliquer pour l'individu i dans le sous-groupe j du groupe k ;

μ , la moyenne globale de la variable à expliquer ;

α_k , l'effet aléatoire du k -ième groupe suivant la loi $\sim \mathcal{N}(0, \sigma_\alpha^2)$;

$\beta(\alpha)_{j(k)}$, l'effet aléatoire emboîté du j -ième sous-groupe dans le k -ième groupe suivant la loi $\sim \mathcal{N}(0, \sigma_\beta^2)$;

$e_{i(jk)}$, l'erreur résiduelle d'effet aléatoire de variance suivant la loi $\sim \mathcal{N}(0, \sigma_\epsilon^2)$ relatif à l'effet de l'arbre i .

Les rapports de variance suivants ont été calculés :

$$\text{Zone :} \quad \sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2)$$

$$\text{Site (Zone) :} \quad \sigma_\beta^2 / (\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2)$$

$$\text{Arbre (Site (Zone)) :} \quad \sigma_\epsilon^2 / (\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2)$$

3 RÉSULTATS

3.1 Variation morphologique des arbres et leur propagules : l'exemple du Mali

3.1.1 Variation générale des mesures morphologiques

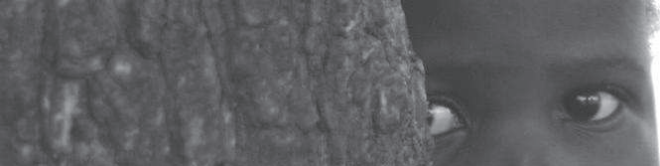
Un total de 254 arbres du Mali a été retenu pour l'analyse morphologique des caractères associés à la taille des arbres et à leurs propagules. Au sein de l'échantillon total les caractères liés à la taille des arbres varient fortement, avec des coefficients de variations entre 20 et 35% (**Tableau III.5**). Dans cette analyse, cependant, le critère de la hauteur des arbres manque pour les arbres des deux sites Nord du Mali (Tori et Sassambourou). Les mesures effectuées sur les fruits des arbres du Mali, soit sur un total de 10160 fruits, montrent des coefficients de variations entre 12 et 35% (**Tableau III.5**). Les caractères associés à la taille des fruits montrent des variations plus faibles que les caractères associés au poids.

Tableau III.5. Statistiques descriptives des mesures morphologiques au Mali

Variable	Unité	N _a	N _f	Min	Max	Moy	CV (%)
DBH	cm	254	-	15.5	115	37.74	35.55
Hteur	m	184	-	5.7	14.5	9.03	19.53
DMH	m	254	-	4.75	17.18	8.67	24.85
LongFruit	cm	254	10160	2.29	5.24	3.26	16.38
LargFruit	cm	254	10160	1.95	3.92	2.77	12.1
PTF	g	254	10160	8.47	42.11	18.68	30
PNF	g	254	10160	2.92	12.61	7.27	24.44
PNS	g	254	10160	1.85	10.36	4.57	35.01

3.1.2 Variation morphologique des peuplements de Karité du Mali

Le **Tableau III.6** montre les statistiques descriptives obtenues pour les mesures morphologiques au sein de chaque site du Mali. S'il on observe que les semenciers du nord du Mali à Tori (DBH moyen = 52,6 cm) et à Sassambourou (DBH moyen = 47,26 cm) ont des diamètres plus importants que ceux observés plus au Sud du pays (entre 30,3 et 35 cm), le patron de variation observé pour les caractères morphologiques des fruits ne semble pas être aussi clair. On note cependant que pour les critères morphologiques des propagules, les



Chapitre III : Variabilité des constituants chimiques des graines

sites de Tori et Sassambourou au Nord-Mali montrent des coefficients de variations légèrement plus faibles que ceux observé au sein des autres sites (**Tableau III.6**).

Tableau III.6. Statistiques descriptives et comparaison des moyennes par site des mesures morphologiques au Mali. Les valeurs en gras sont significatives au seuil 5%. Na : nombre d'arbres échantillonnés ; Nf : nombre de fruits mesurés ; Moy : moyenne exprimée dans l'unité de chaque caractère mesuré ; CV : coefficient de variation exprimé en % ; DBH : diamètre à hauteur de poitrine de l'arbre en cm ; Hteur : hauteur de l'arbre en m ; DMH : diamètre du houppier de l'arbre en m ; LongFruit : longueur du fruit en cm ; LargFruit : largeur du fruit en cm ; PTF : poids du fruit frais en g ; PNF : poids de la noix fraîche en g ; PNS : poids de la noix sèche en g ; Grp : regroupement suggéré par le test de Bonferroni ; d.m. : donnée manquante.

Code Site	Site	N _a	N _f	DBH	Hteur	DMH	LongFruit	LargFruit	PTF	PNF	PNS
				Moy (CV)	Grp	Moy (CV)	Grp	Moy (CV)	Grp	Moy (CV)	Grp
4	Nafégué	44	1760	30.29 (24.9)	A	7.74 (22.9)	A	2.92 (10.8)	B	20.73 (30.1)	B
5	Mperesso	70	2800	34.74 (23)	A	8.52 (19.8)	A	2.67 (13.7)	A	17.09 (35.2)	A
6	Daelan	70	2800	33.26 (31.9)	A	8.38 (19)	A	2.85 (11.1)	B	18.87 (27.8)	AB
7	Tori	35	1400	52.56 (24.7)	B	10.94 (26.1)	B	2.72 (9.3)	A	18.8 (21.6)	AB
8	Sassambourou	35	1400	47.26 (24.1)	B	8.49 (21.1)	A	2.65 (11)	A	18.79 (25.6)	AB
Test de Fisher				31.57	8.18	14.96	3.15	6.44	3	15.85	45.72
p-value				<0.0001	0.0004	<0.0001	0.015	<0.0001	0.019	<0.0001	<0.0001



Différence entre les sites :

Les tests de Fisher des analyses de variances sur caractères morphologiques mesurés au Mali (**Tableau III.6**) sont tous significatif au seuil de 5%. Ainsi, l'ANOVA révèle des différences significatives entre les moyennes de diamètre à hauteur de poitrine (DBH) entre les sites du Mali et le test de Bonferroni permet de classer les sites en deux groupes distincts. Les groupes formés sont cohérents avec une répartition Nord-Sud des sites, avec au Nord les peuplements avec les diamètres les plus importants et au Sud les des diamètres plus petits. La hauteur est également significativement différente malgré l'absence de données pour les deux peuplements échantillonnés au Nord du Mali, le test de Bonferroni suggérant des hauteurs d'arbre significativement plus importantes à l'extrême sud du Mali (Nafégué). Enfin, si le critère de diamètre de houppier montre des différences significatives, le regroupement suggéré par le test de Bonferroni permet uniquement de distinguer le site de Tori des autres.

Les analyses de variances menées sur les caractères associés aux propagules sont toutes significatives. Cependant les tests de Bonferroni effectués ne permettent pas de mettre en évidence de regroupements cohérents ni entre les différents caractères, ni suivant une distribution Nord-Sud. On note cependant que sur l'ensemble des caractères, le site de Nafégué semble montrer les fruits les plus larges et les plus lourds, et Mperesso les propagules les plus petites et les plus légères (**Tableau III.6**).

3.2 La teneur en eau

La distribution des teneurs en eau suit une loi normale (**Figure III.7**) avec un minimum de 1,85 % et un maximum de 9,32%, avec une variance de 0,91, un écart-type de 0,95 et un coefficient de variation de 0,21. La teneur en eau des graines de Karité ne constitue pas dans notre étude un caractère physiologique d'intérêt car les noix ont été séchées après récolte. Cette mesure permet de vérifier que le taux d'humidité des graines est inférieur à 10% et de calculer ensuite le taux de matière grasse en fonction de la quantité de matière « sèche ». Les teneurs en eau inférieures à 10% dans les échantillons permettent d'éviter les activités lipases et les hydrolyses acides des triglycérides en acides gras libres (Verger & Rivière, 1987). Un test de Grubbs a été réalisé au seuil de 5% afin de détecter les échantillons pouvant présenter des valeurs aberrantes. Cinq échantillons sur les 456 ont été détectés

comme extrêmes avec des teneurs en eau comprises entre 7,49% et 9,32%. Aucun défaut particulier n'a été remarqué lors de la réception de ces échantillons. Ces échantillons n'ont pas pour autant été jugés aberrant pour les autres mesures chimiques et leur taux d'humidité relativement élevé est probablement la conséquence de séchage incomplet. Finalement, tous les échantillons ont présenté des teneurs en eau inférieure à 10% comme prévu par le plan d'échantillonnage initial.

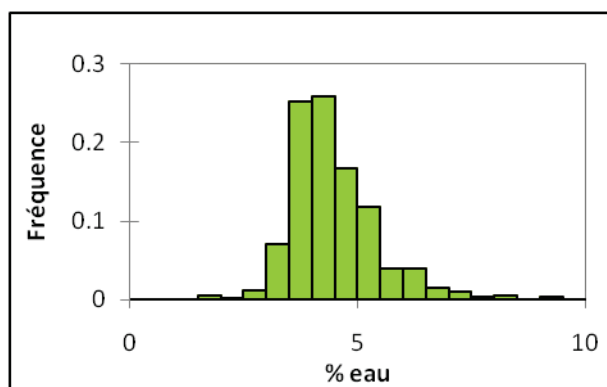


Figure III.7. Histogramme de distribution des teneurs en eau

3.3 La teneur en matière grasse

3.3.1 Variation générale des teneurs en matière grasse

Les teneurs en matière grasse sur l'aire naturelle montre des variations très importantes avec un minimum de 28,01% et un maximum de 61,58% (**Figure III.8**). On observe une moyenne de 49,87% et un coefficient de variation de 10%. Un test de Grubbs au seuil 5%, montre que cinq échantillons sur les 456 présentent des valeurs « aberrantes » avec des valeurs en dessous de 34,3%. Ces taux de matière grasse très faibles sont suspectés dus à des échantillons de fruits récoltés immatures. Ce problème lors de la récolte des fruits du karité est relativement récurrent car la maturité des fruits est difficile à juger. En effet la couleur extérieure des fruits varie entre le vert clair et le marron sans pour autant refléter leur niveau de maturité. Ces cinq échantillons aberrants ont été également détectés comme tel pour les analyses suivantes (profils en acides gras, teneurs en tocophérols et composition de la matière grasse en triglycérides).

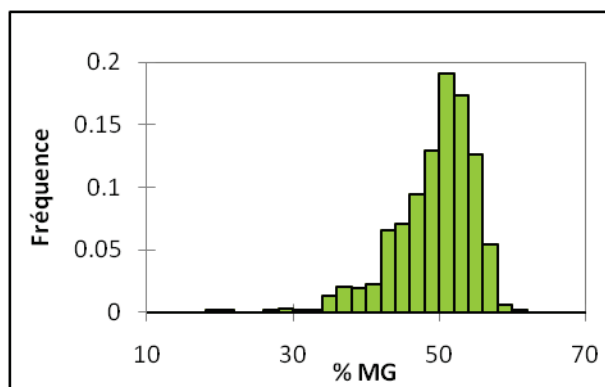


Figure III.8. Histogramme de distribution des teneurs en matière grasse

3.3.2 Caractérisation des régions et populations de Karité pour la matière grasse

Le **Tableau III.7** montre les statistiques descriptives par sites pour les teneurs en matière grasse. Les résultats observés montrent les provenances d'Afrique de l'Est (sous-espèce *nilotica*) ont des teneurs en matière grasse (moyenne: 53,2% ; entre 51,5% et 54,8% en moyenne par site) plus élevées que dans le reste de l'aire naturelle. En Afrique de l'Ouest (sous-espèce *paradoxa*), la teneur moyenne en matière grasse est de 49% et les teneurs moyennes par site sont très variables avec un minimum de 40,3% pour Saraya au Sénégal et un maximum de 52% pour Nafégué au Mali. Les coefficients de variations intra-sites observés sont faibles en Afrique de l'Est (entre 2% et 6%), alors que ceux calculés pour l'Afrique de l'Ouest sont plus importants (entre 5% et 17%). Ce résultat montre que la diversité intra-site des teneurs en matière grasse est plus grande en Afrique de l'Ouest qu'en Afrique de l'Est.

Différence entre régions et sites :

L'analyse de variance en modèle fixe montre que les teneurs moyennes en matière grasse sont très significativement différentes entre les régions Est et Ouest (**Tableau III.8**). Au sein de la région Ouest les moyennes entre sites sont également significativement différentes. Le regroupement effectué par le test de Bonferroni met en évidence l'isolement de Saraya et Kenioto (Sénégal) qui montrent des teneurs en MG significativement plus faibles que les autres sites d'Afrique de l'ouest.

Tableau III.7. Variation des teneurs en matière grasse par région et par site

zone	Site code	Site	N	Moy en % (CV en %)	Min en %	Max en %
AFRIQUE DE L'OUEST	SÉNÉGAL					
	1	Keniototo	12	42.03 (9.48)	36.23	47.77
	2	Samecoutha	18	47.94 (6.84)	43.18	53.98
	3	Saraya	10	40.28 (17.03)	28.01	51.26
	MALI					
	4	Nafégué	40	52.75 (4.64)	47.43	57.44
	5	Mperesso	37	47.4 (11.04)	33.12	56.01
	6	Daelan	37	51.09 (8.37)	42.55	58.08
	7	Tori	35	48.9 (5.49)	43.35	54.74
	8	Sassambourou	35	48.33 (12.15)	34.18	58.34
	BURKINA FASO					
	9	Titao	18	47.75 (8.87)	37.11	52.78
	10	Guibare	21	49.21 (7.3)	43.51	54.29
	GHANA					
	11	Kulbia	34	48.84 (8.03)	38.63	55.38
	12	Tolon	35	51.42 (7.95)	43.6	59.66
	13	Kawampe	35	49.36 (10.81)	35.58	56.8
REGION OUEST			367	49.07 (10.17)	28.01	59.66
AFRIQUE DE L'EST	UGANDA					
	14	Katakwi	25	54.08 (4.96)	49.4	58.11
	15	Pader	29	52.65 (6.16)	45.76	61.58
	16	Moyo	16	54.8 (2.31)	52.75	57.38
	17	Uleppi-Arua	19	51.53 (4.28)	47.17	56.11
REGION EST			89	53.2 (5.31)	45.76	61.58

En Afrique de l'Est, l'ANOVA révèle des différences significatives avec deux groupes chevauchants. Au sein du transect Nord-Sud au Mali, l'analyse de variance montre des différences significatives et le test de Bonferroni permet d'identifier trois groupes et deux sites extrêmes pourtant géographiquement proches : Nafégué et Mperesso. Enfin, au sein du transect Ghana-Burkina, l'analyse révèle également des différences significatives avec un regroupement en deux groupes.

Analyse des composantes de la variance (modèle aléatoire):

L'estimation des composantes de la variance (**Tableau III.9**) à l'échelle de l'aire naturelle révèle que 31% de la variation des teneurs en matière grasse peut s'expliquer par l'effet « zone » et 23% par l'effet « site » au sein de chaque zone. Cependant les teneurs en matière grasse varient fortement entre les individus au sein de chaque site représentant plus de 45% de la variance.

Chapitre III : Variabilité des constituants chimiques des graines

Au sein de la région Ouest près de 35% de la variance observée est lié à un effet « site ». On peut constater que cet effet « site » est plus faible au sein de l'Afrique de l'Est (20%) qui n'est représentée que par l'Ouganda, avec une variance résiduelle forte de près de 80%.

Tableau III.8. Comparaison des moyennes entre régions, sites et transects pour la teneur en matière grasse. Les valeurs en gras sont significatives au seuil de 5%.

Effet			Moy (en %)	N	Groupe	F p
ZONE (SITE)	AIRE NATURELLE	Ouest	49.07	367	A	56.52
		Est	53.2	89	B	<0.0001
	OUEST	3 Saraya	40.28	10	A	10.75 <0.0001
		1 Kenioto	42.03	12	A	
		5 Mperesso	47.4	37	B	
		9 Titao	47.75	18	BC	
		2 Samecouta	47.94	18	BC	
		8 Sassambourou	48.33	35	BC	
		11 Kulbia	48.84	34	BC	
		7 Tori	48.9	35	BC	
		10 Guibare	49.21	21	BCD	
		13 Kawampe	49.36	35	BCD	
		6 Daelan	51.09	37	CD	
		12 Tolon	51.42	35	CD	
		4 Nafégué	52.75	40	D	
	EST	17 Uleppi-Arua	51.53	19	A	6 0.0009
		15 Pader	52.65	29	AB	
		14 Katakwi	54.08	25	B	
		16 Moyo	54.8	16	B	
TRANSECT (SITE)	TRANSECT MALI	5 Mperesso	47.4	37	A	9.86 <0.0001
		8 Sassambourou	48.33	35	AB	
		7 Tori	48.9	35	AB	
		6 Daelan	51.09	37	BC	
		4 Nafégué	52.75	40	C	
	TRANSECT GHANA- BURKINA	9 Titao	47.75	18	A	2.66 <0.0001
		11 Kulbia	48.84	34	AB	
		10 Guibare	49.21	21	AB	
		13 Kawampe	49.36	35	AB	
		12 Tolon	51.42	35	B	

Au sein des deux transects l'analyse des composantes de la variance, révèle une part résiduelle liée à l'effet « arbre » très forte. Ainsi 19% de la variance est expliquée par l'effet « site » au Mali et seulement 5,5% au sein du transect Ghana-Burkina.

Tableau III.9. Part relative des différentes sources de variation pour la matière grasse selon différentes échelles

Source de variation	Au sein de l'aire naturelle			Au sein de la région Ouest		Au sein de la région Est		Au sein du transect Mali		Au sein du transect Ghana-Burkina	
	Zone	Site	Arbre	Site	Arbre	Site	Arbre	Site	Arbre	Site	Arbre
MG	31.0%	23.2%	45.8%	34.9%	65.1%	20.3%	79.7%	18.9%	81.1%	5.5%	94.5%

3.3.3 Relation entre teneurs en matière grasse et caractères morphologiques

Les coefficients de corrélation de Pearson obtenus par régression linéaire entre les variations morphologiques observées chez les arbres du Mali et les teneurs en matière grasse mesurées dans leurs noix sont résumés dans le **Tableau III.10**. On observe des corrélations très faibles et peu significatives pour tous les caractères. On note cependant que le critère le plus associé à la teneur de la graine en matière grasse est le poids de la noix sèche (PNS).

Tableau III.10. Corrélation entre la teneur en matière grasse et les caractères morphologiques des arbres et des fruits. Les valeurs en gras sont significatives au seuil 5%. MG : teneur en matière grasse ; DBH : diamètre à hauteur de poitrine de l'arbre; Hteur : hauteur de l'arbre; DMH : diamètre du houppier ; LongFruit : longueur du fruit; LargFruit : largeur du fruit; PTF : poids du fruit frais; PNF : poids de la noix fraîche ; PNS : poids de la noix sèche.

Caractère morphologique	DBH	Hteur	HGB	DMH	LongFruit	LargFruit	PTF	PNF	PNS
MG	0.06	0.13	0.13	0	0.07	0.18	0.22	0.22	0.41

3.3.4 Relation avec les facteurs climatiques

L'analyse des corrélations entre la matière grasse et les variables géographiques, altitudinales et climatiques (**Tableau III.11**) au sein de chaque zone montre des valeurs relativement faibles. On note cependant que les résultats significatifs diffèrent fortement entre les deux régions. Notamment les variables liées aux températures semblent positivement corrélées à la matière grasse à l'Est, tandis qu'à l'Ouest les corrélations semblent plutôt négatives. À l'inverse les variables climatiques liées aux précipitations sont plutôt positivement corrélées à la matière grasse en Afrique de l'Ouest et négativement à l'Est. Les corrélations au sein des transect montre semble plus marquées sur le transect

« Mali » que sur le transect « Ghana-Burkina ». Mais là aussi les corrélations restent très faibles et aucun coefficient de Pearson n'est supérieur à 0,5.

Tableau III.11. Analyse de corrélations entre les teneurs en matière grasse et les variables géographiques, topographiques et climatiques. Les valeurs en gras sont significatives au seuil de 5%.

	Région Ouest	Région Est	Transect Mali	Transect Ghana- Burkina
Latitude	-0.18	-0.05	-0.27	-0.15
Longitude	0.24	0.16	-0.25	0.16
Altitude	0.07	-0.3	-0.19	-0.16
DRY_M	-0.07	0.08	-0.13	-0.21
ANN_MEANT	-0.09	0.33	0.15	-0.11
TEMP_RANGE	-0.09	0.23	-0.16	-0.17
ISOTHERM	0.19	-0.05	0.29	0.14
TEMP_SEAS	-0.19	0.16	-0.29	-0.15
MAX_T	-0.23	0.28	-0.06	-0.15
MIN_T	0.06	0.33	0.33	0.17
ANN_T_RANGE	-0.13	0.24	-0.24	-0.16
MEAN_T_WtQ	-0.14	0.42	-0.07	-0.18
MEAN_T_DQ	0	0.31	0.35	0.18
MEAN_T_WaQ	-0.25	0.3	-0.07	-0.14
MEAN_T_CQ	0.02	0.36	0.34	0.14
ANN_PRC	0.03	-0.31	0.24	0.16
PRC_WtM	-0.06	-0.36	0.29	0.08
PRC_DM	0.25	0	0.34	0.13
PRC_SEAS	-0.17	-0.04	-0.21	-0.16
PRC_WtQ	-0.11	-0.31	0.25	0.12
PRC_DQ	0.24	-0.17	0.34	0.12
PRC_WaQ	0.34	-0.21	0.4	0.03
PRC_CQ	0.11	-0.32	0.34	0.13

3.3.5 Variation des teneurs en matière grasse en fonction du type de sol

Une analyse de variance en modèle fixe a été effectuée afin de déterminer l'impact d'un effet « type de sol » dans la variation des teneurs en matière grasse au sein des régions étudiées (**Tableau III.12**). L'analyse de variance est très significative au sein des deux régions Est ($p = 0.0019$) et Ouest ($p < 0.0001$). En Afrique orientale, le test de Bonferroni confirme que les teneurs en matière grasse sont significativement plus faibles chez les individus évoluant sur le sol le moins fertile (site d'Uleppi-Arua). Cette tendance est également

confirmée en Afrique de l'Ouest où l'on observe des teneurs en matière grasse significativement plus faibles sur les Leptosols, concentrant les sites Sénégalais. Aucune différence cependant n'est faite entre les autres types de sols.

Tableau III.12. Comparaison des teneurs moyennes en matière grasse des différents types de sols. Les valeurs en gras sont significatives au seuil 5%.

	Type de sol	Indice de Fertilité	Code des sites Concernés (Pays)	N	MG (%)	Grp	F p
RÉGION EST	Leptosols	1	17 (Ouganda)	19	51.53	A	6.72 0.0019
	Ferralsols	2	14 ; 15 (Ouganda)	54	53.31	B	
	Vertisols	5	16 (Ouganda)	16	54.8	B	
RÉGION OUEST	Leptosols	1	1 ; 2 ; 3 (Sénégal)	40	44.25	A	35.13 <0.0001
	Lixisols	4	5 ; 6 ; 7 ; 8 (Mali) ; 9 ; 10 (Burkina) ; 11 ; 12 ; 13 (Ghana)	287	49.23	B	
	Acrisols	3	4 (Mali)	40	51.75	B	

3.4 Les profils en Acides Gras

3.4.1 Variation générale des compositions en acides gras

Basé sur la lecture des profils chromatographiques des échantillons, sept acides gras, montrant des pourcentages relatifs supérieurs à 0,05% ont été retenus pour l'étude. Nous avons détecté trois acides gras saturés (acide palmitique C16:0, acide stéarique C18:0 et acide arachidique C20:0) ; 2 acides gras cis-mono-insaturés (acide oléique C18:1 n-9 et acide cis-vaccénique C18:1 n-7) ; et 2 acides gras polyinsaturés (linoléique C18:2 n-6 et γ -linoléique C18:3 n-6).

Tableau III.13. Statistiques descriptives pour les profils en acides gras calculé sur l'ensemble de l'échantillon

Variable	Palmitique %	Stéarique %	Oléique %	Vaccénique %	Linoléique %	Linoléique %	Arachidique %	Ratio Sat/Insat
N	455	455	455	455	455	455	455	455
Moy	4.39	38.63	47.95	0.4	7.11	0.3	1.18	0.81
Min	2.71	26.08	38.76	0	3.22	0	0.76	0.46
Max	8.44	48.58	60.04	0.9	11.24	0.69	1.76	1.22
CV (%)	16.89	14.36	10.51	43.62	14.34	53.79	15.77	21.02

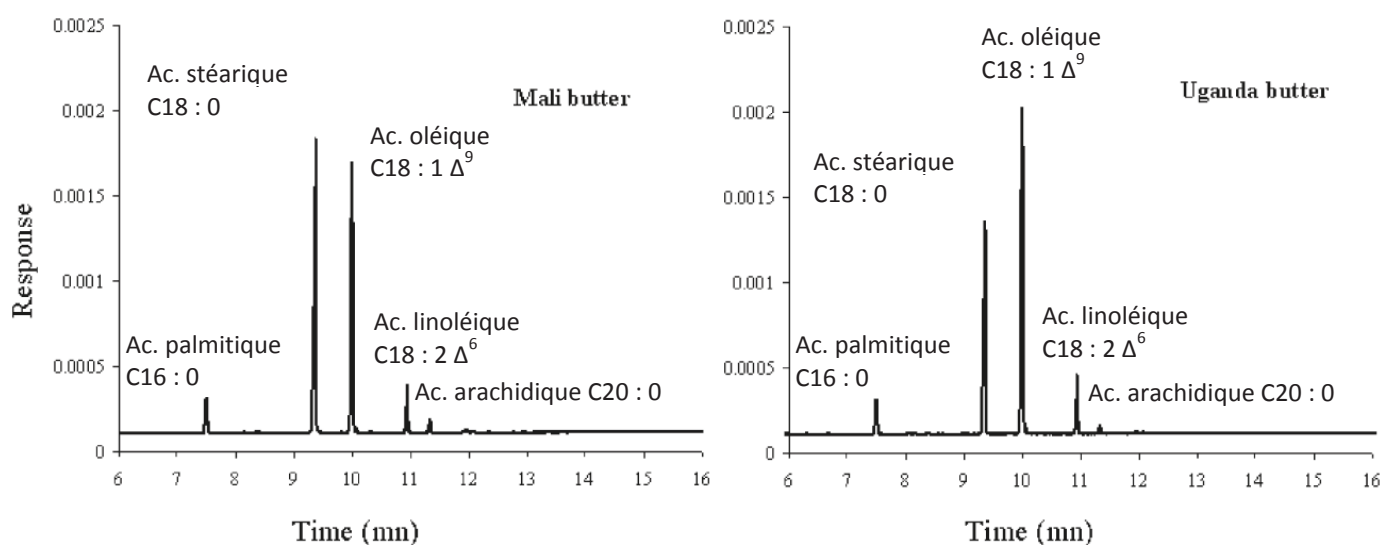
Deux acides gras sont très majoritaires dans le beurre de karité : l'acide stéarique (38,63%) et l'acide oléique (47,95%). Ces deux acides gras représentent à eux seuls 85 à 90% des acides gras totaux contenus dans la graine de karité. L'acide linoléique et l'acide palmitique viennent en second plan avec des pourcentages relatifs respectifs de 7,1 et 4,4%.

Les acides vaccénique (0,4%), linoléique (0,3%) et arachidique (1,2%) sont très minoritaires et peu représentatifs des acides gras de la graine du karité (**Tableau III.13**). Les coefficients de variation sont très variables et liés au type de distribution. Pour les deux acides gras majoritaires de la graine on observe une distribution normale (**Figure III.10**) et les faibles coefficients de variation observés pour ces constituants soulignent une différence réduite entre les individus pour ces deux acides gras.

3.4.2 Caractérisation des populations de Karité pour les profils d'acides gras

Les profils de chromatographie en phase gazeuse de deux échantillons sont présentés à la **Figure III.9**, l'un en provenance du Mali (Afrique de l'Ouest) et l'autre de l'Ouganda (Afrique de l'Est).

Figure III.9. Exemples de chromatogrammes obtenus pour la caractérisation des profils d'acide gras des graines par chromatographie en phase gazeuse.



Ces chromatogrammes confirment la prépondérance des acides stéariques et oléiques dans les profils de d'acides gras des beurres. La comparaison entre les deux profils révèle une proportion d'acide oléique plus marquée dans le beurre ougandais que dans le beurre malien qui montre un taux d'acide stéarique supérieur. Ce résultat est confirmé dans le **Tableau III.14** qui montre que les beurres d'Afrique de l'Est sont plus riches en acide oléique (56,37%) que les beurres d'Afrique de l'Ouest (45,9%). Inversement, le pourcentage relatif d'acide stéarique est plus élevé en Afrique occidentale (40,81%) que dans l'Afrique de

l'Est (29,67%). Le ratio entre acide gras saturés et acide gras insaturés est plus élevé dans la partie Ouest (0,87) par rapport à la zone orientale (0,55) de l'aire naturelle.

On note que les coefficients de variation intrasites pour les deux acides gras majoritaires sont relativement faibles intrasites (entre 2 et 10%). Cependant, on note à l'instar de la matière grasse que les coefficients de variation sont plus élevés en Afrique de l'Ouest qu'en Afrique de l'Est, soulignant une plus grande diversité en Afrique occidentale. Cette observation est également vraie pour les autres acides gras ainsi que pour le ratio calculé entre les acides gras saturés et insaturés.

Tableau III.14. Pourcentages des différents acides gras et coefficient de variation pour chaque site et chaque région

	Site			ac. stéar.	ac. olé.
Zone	code	Site	N	Moy (CV)	Moy (CV)
AFRIQUE DE L'OUEST	SÉNÉGAL				
	1	Keniototo	12	37.51 (7.86)	47.7 (4.22)
	2	Samecouta	18	42.26 (6.33)	44.3 (4.43)
	3	Saraya	10	39.68 (6.65)	45.74 (4.25)
	MALI				
	4	Nafégué	40	41.08 (7.1)	46.6 (5.96)
	5	Mperesso	37	36.75 (9.13)	48.84 (5.6)
	6	Daelan	37	41.81 (10.75)	44.98 (8.9)
	7	Tori	35	40.44 (6.58)	46.87 (4.81)
	8	Sassambourou	35	40.17 (8.35)	46.75 (5.54)
	BURKINA FASO				
	9	Titao	18	41.35 (7.2)	45.46 (5.53)
	10	Guibare	21	41.72 (6.52)	45.51 (5.64)
	GHANA				
	11	Kulbia	34	42.73 (5.51)	44.26 (4.77)
	12	Tolon	35	42.69 (7.17)	44.28 (6.29)
	13	Kawampe	35	40.89 (9.11)	45.06 (6.61)
REGION OUEST			367	40.81 (8.84)	45.9 (6.61)
AFRIQUE DE L'EST	OUGANDA				
	14	Katakwi	25	29.06 (6.67)	57.16 (2.67)
	15	Pader	29	29.84 (6.26)	56.29 (2.88)
	16	Moyo	16	28.57 (4.54)	57.49 (1.86)
	17	Uleppi-Arua	19	31.13 (5.87)	54.53 (3.53)
REGION EST			89	29.67 (6.65)	56.37 (3.35)

Analyse de la corrélation entre les acides gras :

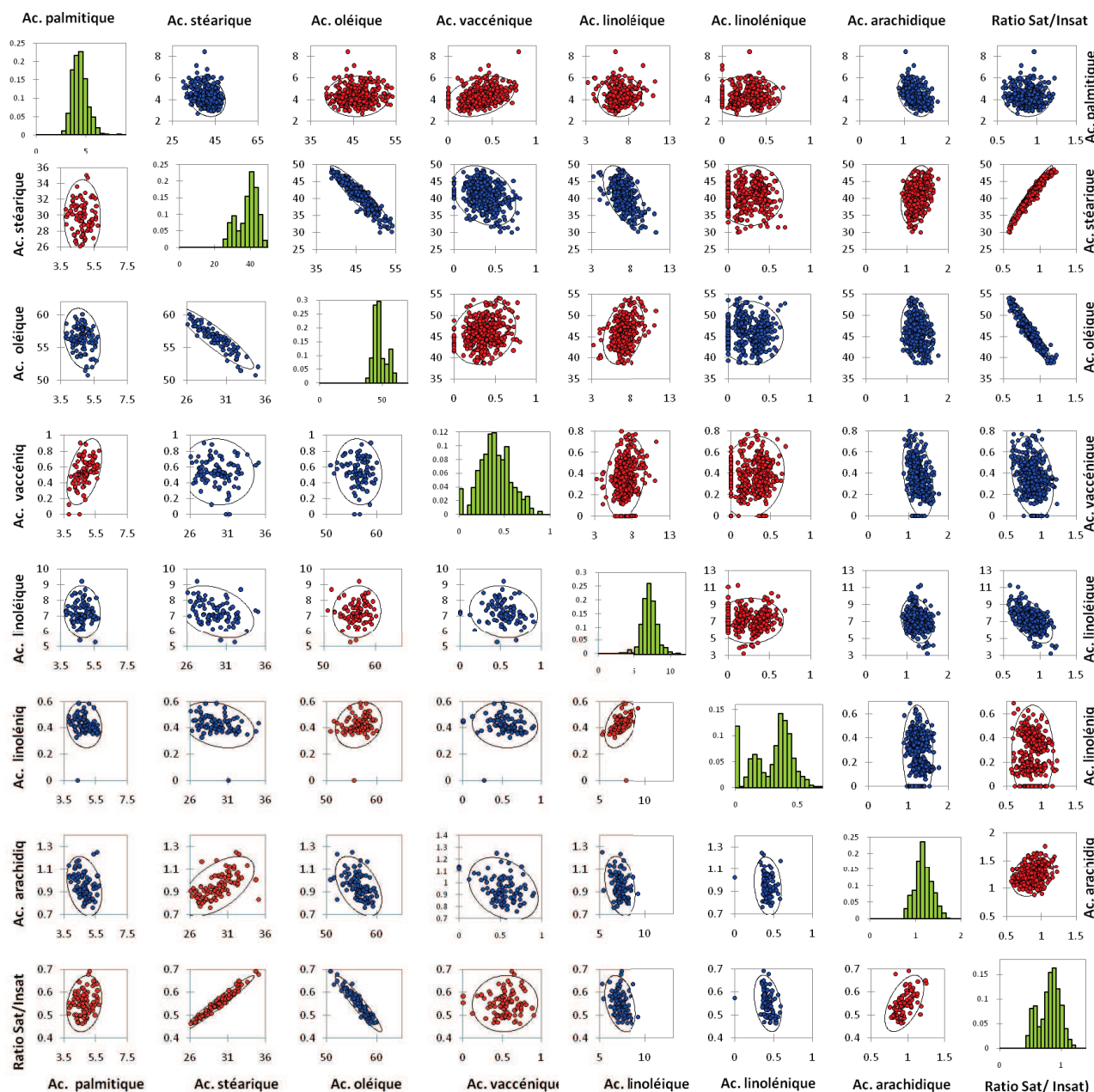


Figure III.10. Histogrammes (échantillon global) et corrélations entre les différents acides gras en Afrique de l'Ouest (moitié supérieure droite de l'image) et en Afrique de l'Est (moitié inférieure gauche de l'image)

Que ce soit au sein de l'Afrique de l'Ouest ou de l'Afrique de l'Est, on observe des corrélations significatives quasiment entre chaque type d'acide gras (**Tableau III.15** et **Figure III.10**). Ces corrélations sont cependant faibles pour la plupart avec des coefficients de corrélation de Pearson (r) inférieurs à 0,6. Néanmoins, on observe des corrélations négatives très fortes et très significatives entre les acides oléique et stéarique avec des coefficients de corrélation à l'Est de 0,91 et à l'Ouest de 0,93. Ce tableau de corrélation ne permet pas de

mettre en évidence de différences marquées entre les corrélations observées à l'Est et à l'Ouest de l'aire naturelle.

Tableau III.15. Corrélation entre les acides gras selon les deux grandes régions. Les coefficients de corrélations au dessus de la diagonale sont pour l'Afrique de l'Ouest, et les autres sont pour l'Afrique de l'Est. Les valeurs en gras sont les coefficients de Pearson significatifs au seuil 5%

Variables	ac. palm.	ac. stéar.	ac. olé	ac. vacc.	ac. linolé.	ac. linolén.	ac. arach.	Sat/Insat
ac. palm.	-	-0.24	0.02	0.34	0.07	0.02	-0.39	-0.04
ac. stéar.	0	-	-0.93	-0.32	-0.54	0.01	0.29	0.97
ac. olé.	-0.23	-0.91	-	0.19	0.31	-0.07	-0.23	-0.95
ac. vacc.	0.46	-0.06	-0.08	-	0.22	0.16	-0.35	-0.27
ac. linolé.	-0.05	-0.35	0.06	-0.13	-	0	-0.21	-0.55
ac. linolén.	-0.13	-0.29	0.2	-0.07	0.41	-	-0.05	0.01
ac. arach.	-0.28	0.57	-0.47	-0.31	-0.31	-0.15	-	0.28
Sat/Insat	0.2	0.98	-0.94	0.04	-0.36	-0.32	0.54	-

Différences entre régions et sites:

Les analyses de variance en modèle fixe pour les deux acides gras majoritaires sont présentées dans le **Tableau III.16**. Les résultats obtenus pour les autres acides gras et le ratio acides gras saturés/insaturés sont présentés dans l'**Annexe III.2**.

Toutes les analyses, sauf pour l'acide linoléique, montrent des effets zones très significatifs ($p < 0,0001$) avec des regroupements de Bonferroni significatifs distinguant les deux régions Est et Ouest. Cette différence est particulière marquée pour l'acide stéarique ($F = 790,8$; $p < 0,0001$) et l'acide oléique ($F = 967,4$; $p < 0,0001$).

Au sein de la région Ouest les différences de moyennes entre sites sont toutes significatives ($p < 0,0001$). Les regroupements obtenus grâce au test de Bonferroni mettent en évidence pour l'acide stéarique et l'acide oléique trois groupes au sein de l'Afrique de l'Ouest (**Tableau III.16**). Ainsi, pour l'acide stéarique et l'acide oléique, le site de Mperesso au Mali est significativement différent des deux sites ghanéens de Tolon et Kulbia. On note pour l'acide palmitique (**Annexe III.2**) que les sites sénégalais (Samecoutha, Saraya et Kenioto) montrent des moyennes significativement plus fortes que les autres sites d'Afrique de l'Ouest.

Chapitre III : Variabilité des constituants chimiques des graines

Dans la région Est, les analyses révèlent des différences significatives pour le ratio saturés/insaturés ($p < 0,0001$) et pour les proportions relatives en acide palmitique ($p = 0,0005$), stéarique ($p = 0,0002$), oléique ($p < 0,0001$) et vaccénique ($p = 0,0116$) dans une moindre mesure (**Tableau III.16** et **Annexe III.2**). Les tests de Bonferroni pour ces constituants révèlent des regroupements similaires avec notamment une distinction systématique des sites de Katakwi (le site le plus au Sud de l'Ouganda) et Uleppi-Arua (le site le plus au Nord de l'Ouganda).

Tableau III.16. Comparaison des moyennes entre régions, sites et transects pour les deux principaux acides gras. Les valeurs en gras sont significatives au seuil de 5%.

EFFET		Acide Stéarique					Acide Oléique				
		Moy	N	Grp	F <i>p</i>	Moy	N	Grp	F <i>p</i>		
ZONE	AIRE	Est	29.67	89	A	790.8 <0.0001	Est	45.9	366	A	967.4 <0.0001
	NATURELLE	Ouest	40.81	366	B		Ouest	56.37	89	B	
ZONE (SITE)	OUEST	5 Mperesso	36.75	37	A	9.1 <0.0001	11 Kulbia	44.26	34	A	8.1 <0.0001
		1 Kenioto	37.51	12	AB		12 Tolon	44.28	35	A	
		3 Saraya	39.68	10	ABC		2 Samecouta	44.3	18	AB	
		8 Sassambourou	40.17	35	BC		6 Daelan	44.98	37	AB	
		7 Tori	40.44	35	BC		13 Kawampe	45.06	35	AB	
		13 Kawampe	40.89	35	BC		9 Titao	45.46	18	AB	
		4 Nafégué	41.08	40	BC		10 Guibare	45.51	21	AB	
		9 Titao	41.35	18	BC		3 Saraya	45.74	10	ABC	
		10 Guibare	41.72	21	C		4 Nafégué	46.6	40	BC	
		6 Daelan	41.81	37	C		8 Sassambourou	46.75	35	BC	
		2 Samecouta	42.26	18	C		7 Tori	46.87	35	BC	
		12 Tolon	42.69	35	C		1 Kenioto	47.7	12	BC	
		11 Kulbia	42.73	34	C		5 Mperesso	48.84	37	C	
EST	16 Moyo	28.57	16	A	7.2 0.0002	17 Uleppi-Arua	54.53	19	A	13.4 <0.0001	
	14 Katakwi	29.06	25	A		15 Pader	56.29	29	B		
	15 Pader	29.84	29	AB		14 Katakwi	57.16	25	B		
	17 Uleppi-Arua	31.13	19	B		16 Moyo	57.49	16	B		
TRANSECT (SITE)	TRANSECT MALI	5 Mperesso	36.75	37	A	12.1 <0.0001	6 Daelan	44.98	37	A	8.1 <0.0001
		8 Sassambourou	40.17	35	B		4 Nafégué	46.6	40	A	
		7 Tori	40.44	35	B		8 Sassambourou	46.75	35	A	
		4 Nafégué	41.08	40	B		7 Tori	46.87	35	A	
		6 Daelan	41.81	37	B		5 Mperesso	48.84	37	B	
	TRANSECT GHANA-BURKINA	13 Kawampe	40.89	35	A	2.33 0.059	11 Kulbia	44.26	34	A	1.44 0.223
		9 Titao	41.35	18	A		12 Tolon	44.28	35	A	
		10 Guibare	41.72	21	A		13 Kawampe	45.06	35	A	
		12 Tolon	42.69	35	A		9 Titao	45.46	18	A	
		11 Kulbia	42.73	34	A		10 Guibare	45.51	21	A	



Chapitre III : Variabilité des constituants chimiques des graines

L'analyse de variance au sein du transect Mali souligne des différences fortement significatives pour toutes les variables sauf pour l'acide linoléique (**Tableau III.16** et **Annexe III.2**). Les regroupements effectués grâce au test de Bonferroni révèlent clairement une distinction de Mperesso par rapport aux autres sites. Ces différences sont très robustes statistiquement car le transect se base sur un échantillonnage important de 184 arbres répartis de façon homogène dans les cinq sites échantillonnés.

Au sein du second transect étudié (Ghana-Burkina) l'analyse de variance présente des différences non significatives pour les deux acides gras majoritaires ainsi que pour le ratio saturés/insaturés (**Tableau III.16** et **Annexe III.2**). De même les tests de Bonferroni plus stringents ne permettent pas de mettre en évidence des différences significatives entre sites pour ces constituants. On note cependant pour l'acide palmitique des différences significatives et un regroupement cohérent avec un gradient de type Nord (Guibaré) vers Sud (Kawampe).

Analyse des composantes de la variance (modèle aléatoire):

L'analyse des composantes de la variance grâce à l'ANOVA en modèle aléatoire (**Tableau III.17**) révèle que l'effet « zone » au sein de l'aire naturelle explique plus de 80% de la variance observée pour les deux acides gras majoritaires et pour le ratio saturés/insaturés, seulement 2 à 4% de la variance est expliqué par un effet « site », et uniquement 10 à 15% de variance résiduelle est observée.

Au sein des régions, on constate que l'effet « site » expliquant la variance des acides oléique et stéarique est bien plus important à l'Est (39,1% pour l'acide oléique et 24,5% pour l'acide stéarique), qu'à l'Ouest de l'aire où l'effet « site » explique respectivement 18,9 et 22,2% de la variance de ces deux acides gras (**Tableau III.17**). En comparant les résultats obtenus au sein des deux régions, on constate que pour les acides gras minoritaires l'effet « site » est bien plus important au sein de la région Ouest (entre 10 et 57% de variance expliquée) qu'au sein de la région Est (entre 0 et 20% de variance expliquée).

Au sein des transects étudiés on observe que les effets « sites » sont bien plus importants au sein du transect Mali avec notamment 23,1% de variance de l'acide stéarique expliquée contre seulement 4,4% au sein du transect Ghana-Burkina (**Tableau III.17**). On

note cependant globalement que les variances résiduelles sont très fortes au sein des deux transects et pour l'ensemble des variables. Ceci suggère donc une part très forte de la variation entre les individus au sein même des sites composant les transect étudiés.

Tableau III.17. Part relative des différentes sources de variation pour les différents acides gras et selon les différentes échelles

Source de Variation	Au sein de l'aire naturelle			Au sein de la région Ouest		Au sein de la région Est		Au sein du transect Mali		Au sein du transect Ghana-Burkina	
	Zone	Site	Arbre	Site	Arbre	Site	Arbre	Site	Arbre	Site	Arbre
Palmitique	1.0	54.0	45.0	56.9	43.1	20.1	79.9	19.3	80.7	13.4	86.6
Stéarique	84.2	3.5	12.3	22.2	77.8	24.5	75.5	23.1	76.9	4.4	95.6
Oléique	87.0	2.7	10.3	18.9	81.1	39.1	60.9	16.0	84.0	1.5	98.5
Linoléique	34.5	11.3	54.2	18.9	81.1	11.8	88.2	19.9	80.1	13.1	86.9
Linoléénique	0.0	9.5	90.5	10.6	89.4	0.0	100.0	12.4	87.6	8.4	91.6
Arachidique	25.2	25.8	49.0	36.2	63.8	0.0	100.0	3.1	96.9	0.0	100.0
Vaccénique	63.2	9.0	27.8	27.1	72.9	0.6	99.4	20.9	79.1	28.0	72.0
Saturé/Insaturé	80.4	4.0	15.6	20.3	79.7	28.8	71.2	22.3	77.7	4.0	96.0

3.4.3 Relation avec les caractères morphologiques

De même que pour les corrélations entre les caractères morphologiques et les teneurs en MG, les coefficients de corrélation de Pearson obtenus entre les variations morphologiques et les pourcentages relatifs en acides gras sont très faibles et peu significatifs (**Tableau III.18**). Cependant à l'instar de la MG, bien que les coefficients de corrélations restent très faibles, le critère morphologique le plus associé à la composition en acide gras des graines est le poids de la noix sèche (PNS).

Tableau III.18. Corrélations entre les acides gras et les critères morphologiques observés. Les valeurs en gras sont significatives au seuil de 5%.

	Palmitique	Stéarique	Oléique	Vaccénique	Linoléique	Linoléénique	Arachidique	Sat/Insat
DBH	0.08	0.06	-0.05	0.06	-0.05	-0.01	-0.13	0.06
Hteur	-0.11	0.15	-0.14	0.02	-0.04	-0.08	0.11	0.15
HGB	-0.09	0.12	-0.07	-0.04	-0.14	-0.02	-0.05	0.11
DMH	0.1	-0.01	0.02	0.09	0.01	0	-0.1	-0.02
Longfruit	-0.05	-0.01	0.03	-0.07	-0.04	-0.01	0.01	-0.02
LargFruit	-0.14	0.14	-0.08	-0.09	-0.2	-0.1	0.08	0.12
Poids	-0.16	0.13	-0.06	-0.09	-0.18	-0.07	0.03	0.1
PNF	-0.32	0.13	-0.02	-0.26	-0.16	-0.02	0.17	0.07
PNS	-0.43	0.32	-0.2	-0.32	-0.24	-0.07	0.23	0.25

3.4.4 Relation avec les facteurs climatiques

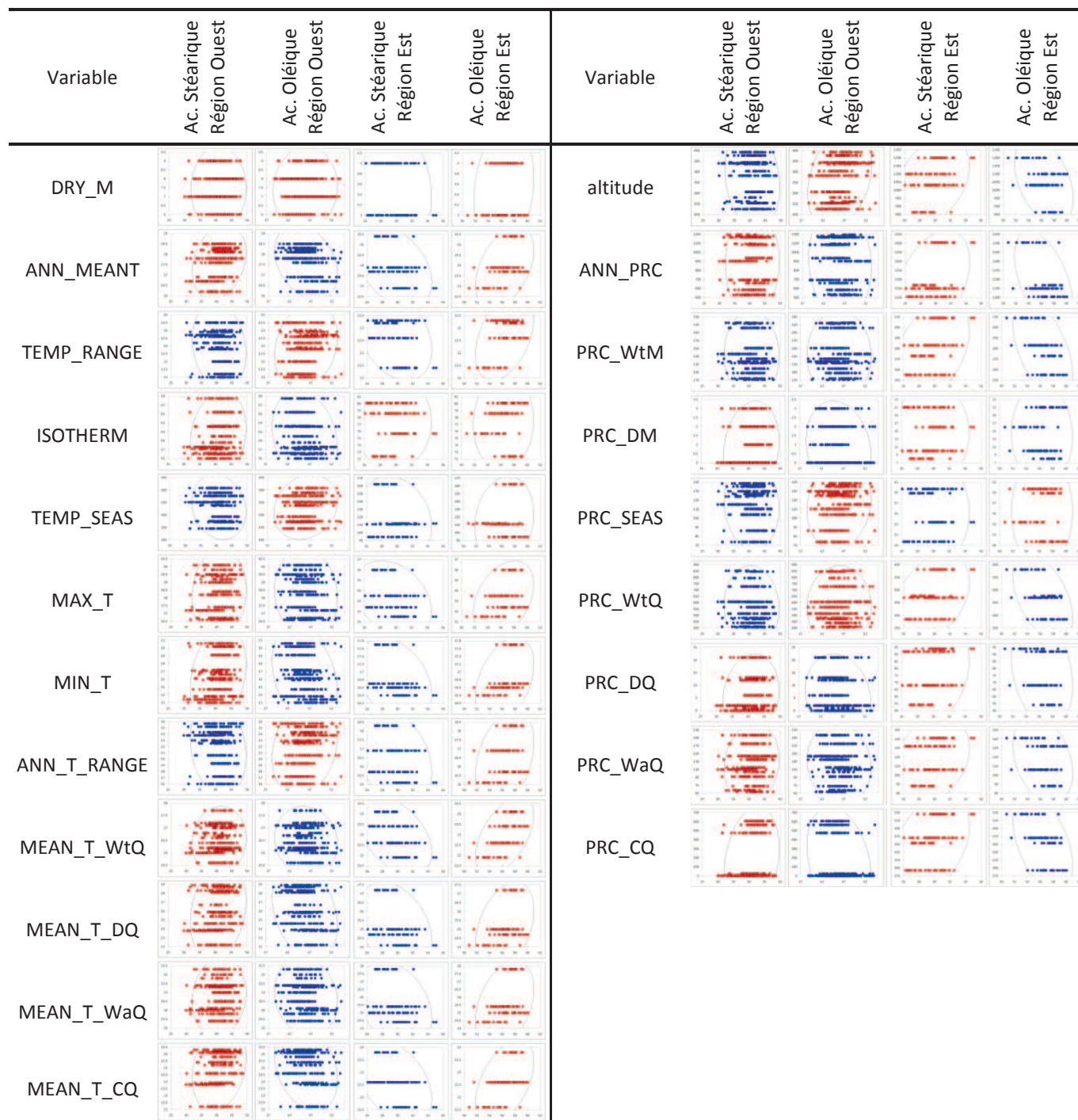
Tableau III.19. Analyse de corrélation entre les acides gras majoritaires et les variables bioclimatiques. Les valeurs en gras sont significatives au seuil de 5%.

Variable	Acide stéarique				Acide oléique				Ratio Sat/Insat			
	Ouest	Est	Transect Mali	Transect Ghana-Burkina	Ouest	Est	Transect Mali	Transect Ghana-Burkina	Ouest	Est	Transect Mali	Transect Ghana-Burkina
Latitude	-0.13	0.06	0.02	0.01	0.17	-0.12	-0.07	0.1	-0.12	0.11	0.07	-0.04
Longitude	0.19	-0.22	0.04	0.21	-0.14	0.33	-0.03	-0.19	0.1	-0.31	0.05	0.22
altitude	-0.18	0.35	-0.11	-0.01	0.3	-0.42	0.18	0.11	-0.26	0.36	-0.18	-0.06
DRY_M	0.01	-0.14	0.16	0.02	0.06	0.18	-0.16	0.07	-0.01	-0.17	0.19	-0.01
ANN_MEANT	0.26	-0.37	0.22	0.11	-0.32	0.44	-0.22	0	0.31	-0.37	0.25	0.07
TEMP_RANGE	-0.17	-0.32	0.04	-0.03	0.22	0.42	-0.15	0.13	-0.16	-0.38	0.14	-0.07
ISOTHERM	0.18	0.01	0.04	-0.06	-0.22	0.04	0.04	-0.06	0.15	-0.06	-0.03	-0.01
TEMP_SEAS	-0.16	-0.16	-0.02	0	0.18	0.15	-0.04	0.1	-0.13	-0.11	0.04	-0.05
MAX_T	0.08	-0.32	0.21	0.04	-0.09	0.37	-0.24	0.07	0.15	-0.32	0.27	0
MIN_T	0.26	-0.34	0.19	0.03	-0.31	0.39	-0.07	-0.13	0.26	-0.32	0.1	0.07
ANN_T_RANGE	-0.18	-0.29	0	0	0.22	0.35	-0.1	0.11	-0.16	-0.3	0.09	-0.05
MEAN_T_WtQ	0.07	-0.44	0.16	-0.13	-0.09	0.53	-0.19	0.17	0.1	-0.45	0.21	-0.16
MEAN_T_DQ	0.23	-0.33	0.16	0.09	-0.3	0.38	-0.1	-0.16	0.27	-0.32	0.12	0.13
MEAN_T_WaQ	0.1	-0.32	0.15	0.06	-0.14	0.37	-0.19	0.06	0.18	-0.31	0.21	0.01
MEAN_T_CQ	0.23	-0.41	0.15	0.1	-0.29	0.49	-0.1	-0.14	0.27	-0.42	0.12	0.14
ANN_PRC	0.02	0.38	-0.05	0.01	-0.09	-0.51	0.09	-0.11	0.06	0.45	-0.1	0.06
PRC_WtM	-0.02	0.4	0.02	0.17	-0.01	-0.51	0.03	-0.17	0.03	0.44	-0.03	0.2
PRC_DM	0.16	0.04	0.14	-0.06	-0.17	-0.03	-0.03	-0.06	0.12	0.04	0.06	-0.01
PRC_SEAS	-0.09	-0.01	0.09	0.03	0.15	0.01	-0.12	0.08	-0.08	-0.03	0.14	-0.01
PRC_WtQ	-0.05	0.36	-0.05	0.21	0.01	-0.48	0.08	-0.19	0.01	0.42	-0.09	0.23
PRC_DQ	0.14	0.22	0.12	-0.08	-0.16	-0.26	-0.03	-0.05	0.1	0.23	0.05	-0.03
PRC_WaQ	0.12	0.26	0.23	-0.08	-0.1	-0.32	-0.16	-0.06	0.06	0.28	0.19	-0.02
PRC_CQ	0.24	0.36	0.13	0.12	-0.29	-0.47	-0.03	-0.18	0.24	0.41	0.05	0.16

L'analyse de corrélation par régression linéaire entre les acides gras et les variables géographiques altitudinales et climatiques au sein de chaque région Est et Ouest montre des relations faibles avec des coefficients de corrélation de Pearson (r) inférieurs à 0,5 (**Tableaux III.19 et III.20**). Il est cependant intéressant de remarquer que les acides oléique et stéarique sont systématiquement inversement corrélés pour chaque variable. Plus encore, on peut observer que les variables associées aux températures (ANN_MEAN_T ; MIN_T ; MEAN_T_DQ ; MEAN_T_WaQ ; MEAN_T_CQ) sont systématiquement inversement corrélées entre l'Est et l'Ouest pour les deux acides gras majoritaires et le ratio saturés/insaturés. Ceci suggère que l'impact des facteurs climatiques est très différent entre l'Est et l'Ouest.

Chapitre III : Variabilité des constituants chimiques des graines

Tableau III.20. Graphiques XY de corrélation entre les deux acides gras majoritaires et les variables climatiques et topographiques au sein des deux régions Ouest et Est. Les coefficients de Pearson en gras sont significatifs au seuil 5%.



Au sein de l'Afrique de l'Est on constate que les variables liées aux températures sont négativement corrélées à l'acide stéarique et positivement à l'acide oléique, les variables associées à la pluviométrie sont elles positivement corrélées à l'acide stéarique et négativement à l'acide oléique. Ainsi cette observation est moins évidente au sein de l'Afrique de l'Ouest, où il semble plus difficile de mettre à jour des corrélations intéressantes



entre les variables climatiques et les deux acides gras majoritaires. Les corrélations au sein des transects sont très faibles et peu significatives.

3.5 Les Tocophérols

3.5.1 Variation générale des teneurs en tocophérols de la matière grasse

Après lecture des chromatogrammes obtenus en HPLC et intégration des pics de tocophérols en utilisant des gammes d'étalons de concentrations connues, on observe sur les 273 échantillons analysés (**Tableau III.21**) que l' α -tocophérol est largement majoritaire dans la matière grasse des graines de karité avec une teneur moyenne d'environ 112 ppm (parties par million) soit 112 $\mu\text{g/g}$ de beurre et un coefficient de variation relativement important (29%). Le γ -tocophérol vient en deuxième position avec une teneur moyenne d'environ 13 ppm et un coefficient de variation très élevé (77%). Les tocophérols bêta et delta sont présents mais en très faible quantité et avec des coefficients de variations extrêmement importants de plus de 110%. Les résultats obtenus se trouvent dans une gamme inférieure à ceux publiés antérieurement (Maranz & Wiesman, 2004). Ainsi les auteurs avaient observé sur un échantillonnage de 92 échantillons venant de 22 sites, des teneurs moyennes en α -tocophérol identique à celle calculée dans notre étude, mais des teneurs en bêta, gamma et delta tocophérols deux à trois fois plus importantes.

Tableau III.21. Statistiques descriptives pour les teneurs en tocophérols calculées sur l'ensemble de l'échantillon

Variable	Alpha toco. (ppm)	Beta toco. (ppm)	Gamma toco. (ppm)	Delta toco. (ppm)
N	273	273	273	273
Moy	111.73	1.22	12.82	0.49
Min	24	0.01	0.01	0.01
Max	214.8	6.3	51.8	6.4
cv	28.62	111.04	76.73	227.63

3.5.2 Caractérisation des populations de Karité pour les teneurs en tocophérols

Globalement, à l'instar des autres constituants de la matière grasse, on observe des coefficients de variation intrarégionaux et intrasites plus importants à l'Ouest qu'à Est (**Tableau III.22**). Ainsi ceci suggère une fois encore une diversité plus importante en Afrique occidentale. On observe également que la teneur moyenne en α -tocophérol est plus élevée à l'Est (122 $\mu\text{g/g}$) qu'à l'Ouest (111 $\mu\text{g/g}$). C'est aussi le cas pour les autres formes de tocophérols et ce en particulier pour le γ -tocophérol avec une teneur moyenne de 34 $\mu\text{g/g}$ à l'Est et seulement 11 $\mu\text{g/g}$ en Afrique de l'Ouest (**Tableau III.22**).

Tableau III.22. Teneurs en tocophérols et coefficient de variation pour chaque site et chaque région

ZONE	Site code	Site	N	Alpha toco. Moy (CV)	Beta toco. Moy (CV)	Gamma toco. Moy (CV)	Delta toco. Moy (CV)
AFRIQUE DE L'OUEST	SÉNÉGAL						
	1	Keniotto	5	54.85 (52.83)	0.53 (196.24)	14.45 (30.49)	0.01 (0)
	2	Samecouta	9	74.69 (43.59)	0.36 (192.97)	12.62 (81.27)	0.1 (269.32)
	3	Saraya	7	128.03 (44.17)	0.86 (169.99)	12.86 (35.65)	0.01 (0)
	MALI						
	4	Nafégué	40	96.56 (19.94)	1.08 (113.54)	11.73 (75.31)	0.68 (196.35)
	5	Mperesso	37	138.45 (21.39)	1.87 (77.73)	11.84 (48.54)	0.38 (207.61)
	6	Daelan	37	94.31 (26.94)	0.65 (130.84)	6.85 (56.21)	0.01 (0)
	7	Tori	35	112.68 (23.44)	1.05 (85.07)	8.11 (44.91)	0.27 (314.59)
	8	Sassambourou	35	121.42 (24.95)	0.83 (124.14)	12.01 (53.35)	0.38 (194.48)
	BURKINA FASO						
	9	Titao	7	134.44 (17.15)	1.55 (81.73)	14.41 (36.48)	0.01 (0)
	10	Guibare	7	129.99 (14.61)	0.91 (130.24)	9.4 (38.6)	0.17 (248.61)
	GHANA						
	11	Kulbia	9	101.69 (20.86)	0.71 (167.11)	7.62 (42.52)	0.25 (288.16)
	12	Tolon	9	110.88 (14.44)	1.23 (117.94)	11.8 (73.53)	0.66 (157.51)
	13	Kawampe	9	98.75 (24.15)	1.11 (117.26)	12.15 (53.79)	0.88 (89.12)
REGION OUEST			245	110.6 (29.23)	1.06 (112.75)	10.53 (61.38)	0.34 (245.36)
AFRIQUE DE L'EST	OUGANDA						
	14	Katakwi	9	129.4 (13.78)	2.49 (73.98)	35.3 (34.38)	1.25 (87.37)
	15	Pader	4	110.97 (30.31)	1.48 (81.33)	24.07 (18.97)	0.34 (193.17)
	16	Moyo	9	104.13 (19.21)	3.03 (34.08)	42 (12.65)	3.25 (82.62)
	17	Uleppi-Arua	5	143 (11.35)	4.92 (20.4)	41.54 (14.33)	4.76 (24.88)
REGION EST			27	122.03 (22.11)	2.68 (67.31)	33.7 (31.94)	1.89 (110.7)

Au sein de la région Ouest on observe des variations très importantes entre les sites avec par exemple pour l' α -tocophérol un minimum de 55 $\mu\text{g/g}$ pour Kénioto au Sénégal et un maximum de 138 $\mu\text{g/g}$ pour Mperesso au Mali (**Tableau III.22**).

Analyse de la corrélation entre les formes de tocophérols :

On observe des corrélations relativement significatives entre les différentes formes de tocophérol (**Tableau III.23** et **Figure III.11**). Cependant, les corrélations sont faibles, pour la plupart avec des coefficients de corrélation de Pearson (r) inférieurs à 0,7. Néanmoins, on observe des différences marquées entre les corrélations observées à l'Est et à l'Ouest de l'aire naturelle. En effet, elle semble plus forte en Afrique orientale. Il faut cependant relativiser ce résultat du fait d'un déséquilibre important dans l'échantillonnage entre les deux régions.

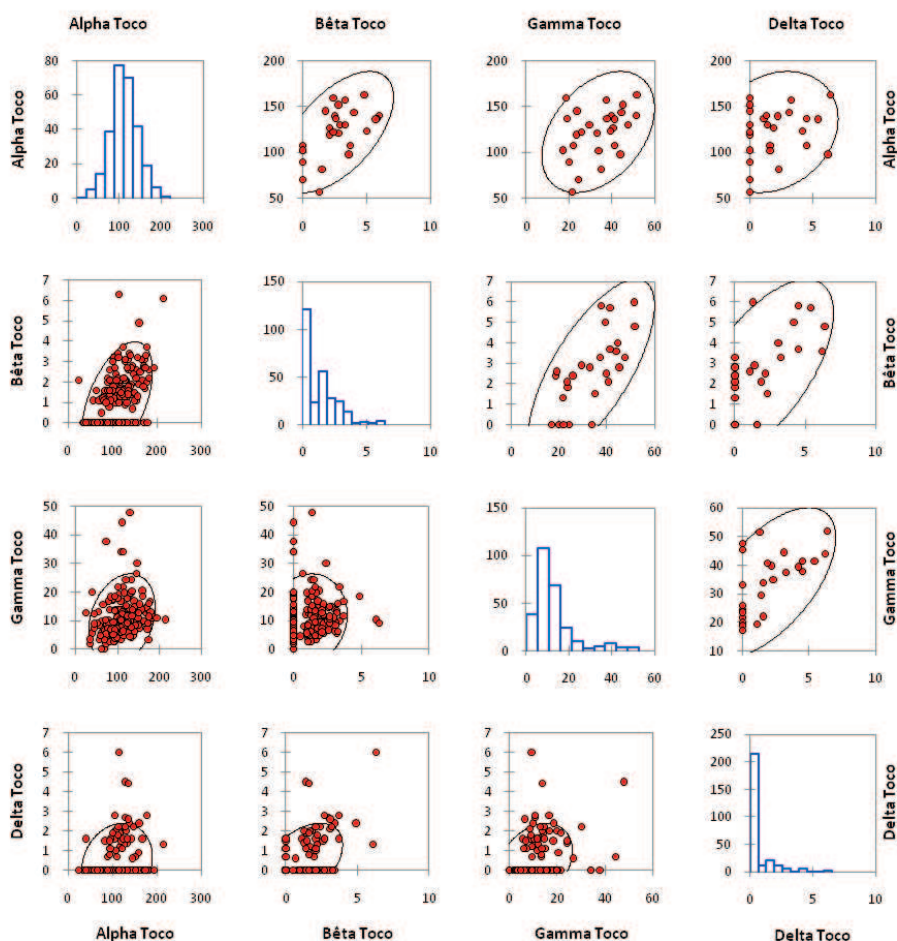


Figure III.11. Histogrammes (échantillon global) et corrélations entre les différentes formes de tocophérols en Afrique de l'Est (moitié supérieure droite de l'image) et en Afrique de l'Ouest (moitié inférieure gauche de l'image)

Tableau III.23. Corrélation entre les formes de tocophérols selon les deux grandes régions. Les coefficients de corrélations au dessus de la diagonale sont pour l'Afrique de l'Est, et les autres sont pour l'Afrique de l'Ouest. Les valeurs en gras sont les coefficients de Pearson significatifs au seuil 5%

Variables	Alpha Toco	Bêta Toco	Gamma Toco	Delta Toco
Alpha Toco	-	0.58	0.39	0.22
Bêta Toco	0.50	-	0.71	0.63
Gamma Toco	0.25	0.10	-	0.61
Delta Toco	0.16	0.46	0.31	-

Différence entre régions et sites:

Les analyses de variance en modèle fixe pour les deux formes α et γ des tocophérols sont présentées dans le **Tableau III.24**. Les résultats obtenus pour le β -tocophérol et le δ -tocophérol sont présentés dans l'**Annexe III.2**.

À l'échelle de l'aire naturelle, l'analyse de variance est non significative pour l' α -tocophérol et très significative pour le γ -tocophérol. Ainsi pour ce dernier, le test de Bonferroni confirme la distinction entre les deux régions Est et Ouest. À contrario, pour l' α -tocophérol aucune différence significative entre les moyennes des deux régions n'est mise en évidence (**Tableau III.24**). L'analyse de variance est significative pour les deux formes minoritaires β et δ des tocophérols, différenciant les teneurs moyennes de l'Est et de l'Ouest (**Annexe III.2**).

Au sein de la région Ouest, l'analyse de variance pour le γ -tocophérol est très faiblement significative (**Tableau III.24**). Cependant le test de Bonferroni par paire de site est non significatif et ne permet pas d'identifier de groupes. À l'inverse, l'analyse de variance au sein de l'Ouest pour l' α -tocophérol est très significative (**Tableau III.24**). Ainsi le test de Bonferroni propose quatre groupes distinguant en particulier les sites de Kenioto (Sénégal), Samecouta (Sénégal), Daelan (Mali) et Nafégué (Mali) d'une part, des sites de Sassambourou (Mali), Guibare (Burkina-Faso), Titao (Burkina-Faso) et Mperesso (Mali).

Au sein de la région orientale, l'analyse de variance est non significative pour l' α -tocophérol. Elle est en revanche significative pour le γ -tocophérol et le test de Bonferroni suggère que le site de Pader montre une teneur moyenne en γ -tocophérol significativement plus faible que celles des sites d'Uleppi-Arua et de Moyo (**Tableau III.24**).

Au sein des transects les analyses de variance sont très significatives pour l' α -tocophérol (**Tableau III.24**). Les tests de Bonferroni permettent en particulier de proposer trois groupes pour le transect Ghana-Burkina avec un regroupement en accord avec la distribution des sites du Sud vers le Nord. Le regroupement proposé pour le transect Mali semble lui moins évident mais permet de

Chapitre III : Variabilité des constituants chimiques des graines

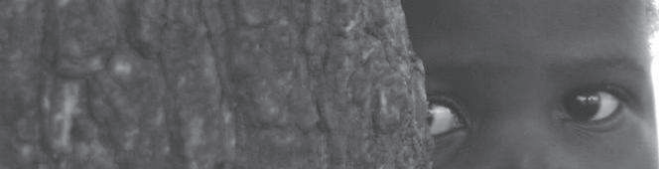
révéler des différences significatives notamment entre Mperesso et Daelan. L'analyse de variance pour le γ -tocophérol est significative pour le transect Mali uniquement (**Tableau III.24**). Le regroupement en deux classes ne semble pas cependant répondre à une distribution géographique car Daelan, qui semble une population extrême du transect, se trouve exactement au milieu de celui-ci.

Tableau III.24. Comparaison des moyennes entre régions et sites pour les deux formes de tocophérols majoritaires. Les valeurs en gras sont significatives au seuil de 5%.

		Alpha toco.					Gamma toco.				
			Moy	N	Grp	F <i>p</i>		Moy	N	Grp	F <i>p</i>
ZONE	AIRE	Ouest	110.6	246	A	3.13	Ouest	11	246	B	267.59
	NATURELLE	Est	122.03	27	A	0.08	Est	34	27	A	<0.0001
ZONE (SITE)	OUEST	1 Kenioto	54.85	5	A	9.74 <0.0001	6 Daelan	6.85	37	A	2.76 0.016
		2 Samecouta	74.69	9	A		11 Kulbia	7.62	9	A	
		6 Daelan	94.31	37	AB		7 Tori	8.11	35	A	
		4 Nafégué	96.56	40	AB		10 Guibare	9.4	7	A	
		13 Kawampe	98.75	9	ABC		4 Nafégué	11.73	40	A	
		11 Kulbia	101.69	9	ABC		12 Tolon	11.8	9	A	
		12 Tolon	110.88	9	ABCD		5 Mperesso	11.84	37	A	
		7 Tori	112.68	35	BCD		8 Sassambourou	12.01	35	A	
		8 Sassambourou	121.42	35	CD		13 Kawampe	12.15	9	A	
		3 Saraya	128.03	7	CD		2 Samecouta	12.62	9	A	
		10 Guibare	129.99	7	CD		3 Saraya	12.86	7	A	
		9 Titao	134.44	7	CD		9 Titao	14.41	7	A	
		5 Mperesso	138.45	37	D		1 Kenioto	14.45	5	A	
	EST	16 Moyo	104.13	9	A	2.81 0.06	15 Pader	24.07	4	A	7.03 0.0016
		15 Pader	110.97	4	A		14 Katakwi	35.3	9	AB	
		14 Katakwi	129.4	9	A		17 Uleppi-Arua	41.54	5	B	
		17 Uleppi-Arua	143	5	A		16 Moyo	42	9	B	
TRANSECT (SITE)	TRANSECT MALI	6 Daelan	94.31	37	A	18.05 <0.0001	6 Daelan	6.85	37	A	5.91 0.0002
		4 Nafégué	96.56	40	AB		7 Tori	8.11	35	AB	
		7 Tori	112.68	35	BC		4 Nafégué	11.73	40	B	
		8 Sassambourou	121.42	35	CD		5 Mperesso	11.84	37	B	
		5 Mperesso	138.45	37	D		8 Sassambourou	12.01	35	B	
	TRANSECT GHANA-BURKINA	13 Kawampe	98.75	9	A	4.85 0.003	13 Kawampe	7.62	9	A	1.55 0.21
		11 Kulbia	101.69	9	AB		10 Guibare	9.4	7	A	
		12 Tolon	110.88	9	ABC		12 Tolon	11.8	9	A	
		10 Guibare	129.99	7	BC		11 Kulbia	12.15	9	A	
		9 Titao	134.44	7	C		9 Titao	14.45	7	A	

Analyse des composantes de la variance (modèle aléatoire):

L'analyse des composantes de la variance par le modèle aléatoire (**Tableau III.25**) révèle que l'effet « zone » au sein de l'aire naturelle explique plus de 84% de la variance



Chapitre III : Variabilité des constituants chimiques des graines

observée pour le γ -tocophérol, avec seulement 12% de la variance est expliqué par un effet arbre (variance résiduelle). À contrario, pour l'autre forme dominante de tocophérol (alpha), seulement 1% de la variance est expliquée par l'effet « zone » avec un effet site sur l'aire naturelle non négligeable (35%).

Au sein de la région Ouest, on note que l'effet « site » dans la variance de la teneur en α -tocophérol est fort avec une contribution de près de 38% (**Tableau III.25**). À l'inverse, on observe pour le γ -tocophérol que la part de l'effet « site » est bien plus faible car inférieur à 10% avec une variance résiduelle très forte soulignant une forte contribution de l'effet « arbre » (90%).

Au sein de la région Est, la contribution de l'effet « site » dans la variabilité des teneurs en α -tocophérol est non négligeable (23%) mais semble bien moins forte qu'en Afrique occidentale (**Tableau III.25**). Inversement, 47% de la variation des teneurs en γ -tocophérol s'explique à l'Est par un effet « site ». Ce résultat tranche avec ce qu'y est observé en Afrique de l'Ouest où la contribution des « sites » est négligeable dans l'expression de la variance du γ -tocophérol (**Tableau III.25**).

Tableau III.25. Part relative des différentes sources de variation pour les différentes formes de tocophérols et selon les différentes échelles

Source de Variation	Au sein de l'aire naturelle			Au sein de la région Ouest		Au sein de la région Est		Au sein du transect Mali		Au sein du Transect Ghana-Burkina	
	Zone	Site	Arbre	Site	Arbre	Site	Arbre	Site	Arbre	Site	Arbre
Alpha toco.	1.3	34.9	63.8	37.9	62.1	22.6	77.4	31.3	68.7	32.3	67.7
Beta toco.	49.3	9.2	41.5	7.6	92.4	46.8	53.2	12.7	87.3	0	100
Gamma toco.	84.2	3.6	12.3	8.4	91.6	47	53	11.8	88.2	6	94
Delta toco.	56.8	20.6	22.6	5.1	94.9	67.8	32.2	5	95	12.1	87.9

3.5.3 Relation avec les caractères morphologiques

Le **Tableau III.26** révèle qu'il n'existe aucune relation forte entre les variations en morphologiques des semenciers et les teneurs en tocophérols. En effet, les coefficients de corrélations de Pearsons sont pour la plupart non-significatifs et toujours très faibles ($-0.4 < r < 0.4$).

Chapitre III : Variabilité des constituants chimiques des graines

Tableau III.26. Corrélation entre teneurs en tocophérols et variations morphologiques. Les valeurs en gras sont significatives au seuil de 5%.

	Alpha	BetaToco	Gamma	Delta
DBH	0.12	0.15	0.13	0.11
Hteur	-0.12	0.06	-0.1	0.02
HGB	-0.03	-0.01	0.04	0
DMH	0.01	0.05	0	0
Longfruit	-0.14	-0.13	-0.04	0.01
LargFruit	-0.23	-0.09	-0.07	0.09
Poids	-0.16	-0.07	0.05	0.09
PNF	-0.24	-0.15	-0.01	0.01
PNS	-0.38	-0.15	-0.02	0.01

3.5.4 Relation avec les facteurs climatiques

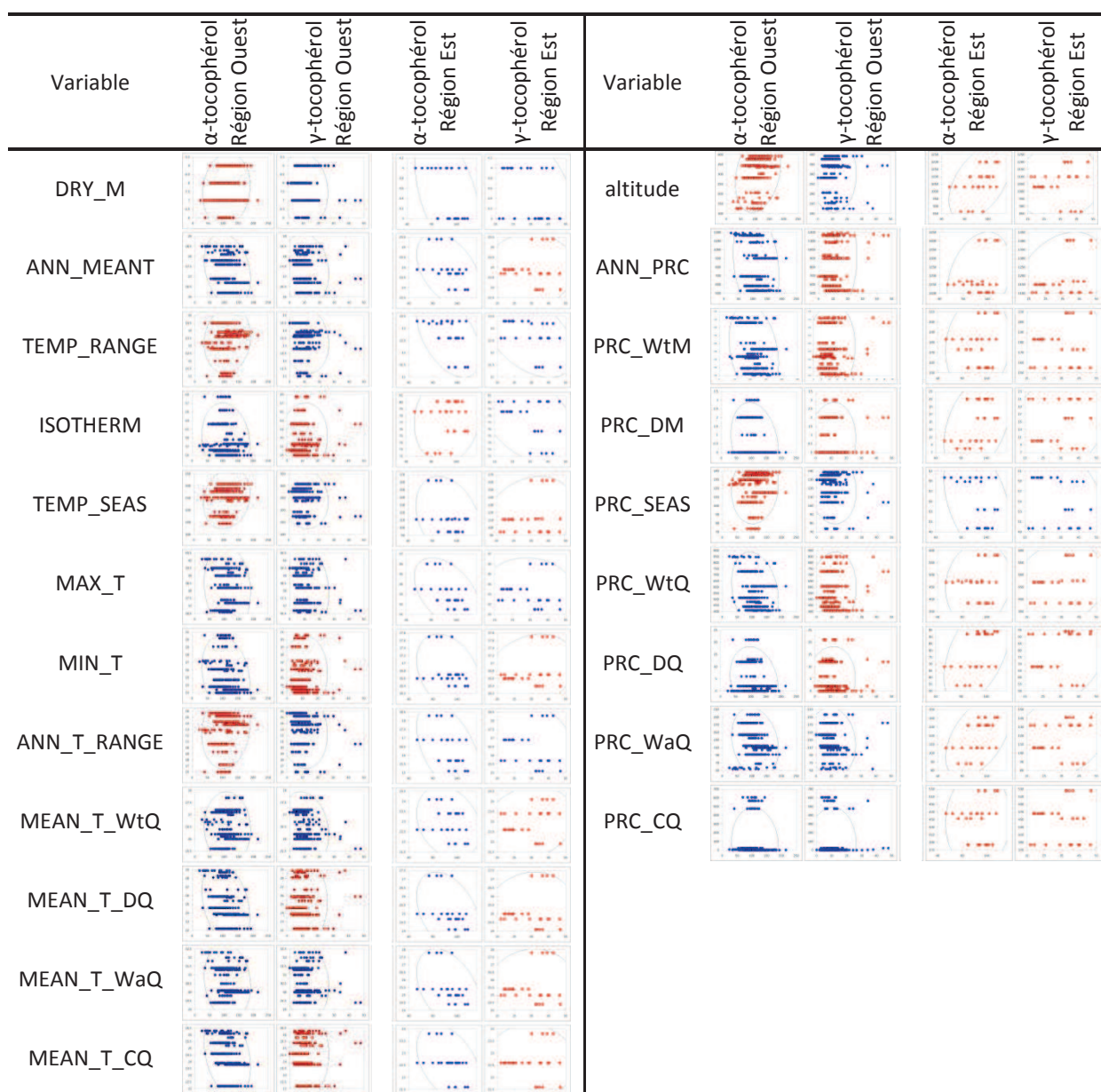
Tableau III.27. Corrélations entre tocophérols et variables environnementales. Les valeurs en gras sont significatives au seuil de 5%.

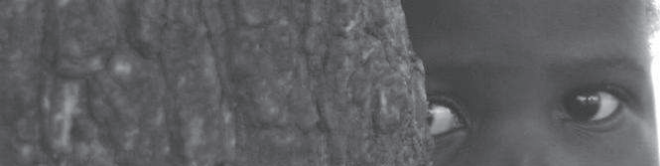
	Alpha				Gamma			
	Ouest	Est	Transect Mali	Transect Ghana-Burkina	Ouest	Est	Transect Mali	Transect Ghana-Burkina
Y	0.16	-0.22	0.15	0.52	-0.09	0.12	-0.12	0.01
X	0.2	-0.1	0.15	-0.39	-0.1	-0.41	-0.07	-0.26
altitude	0.27	0.48	0.25	0.53	-0.03	0.07	0.18	0.03
DRY_M	0.11	-0.48	0	0.39	-0.09	-0.38	-0.07	-0.04
ANN_MEANT	-0.25	-0.43	-0.31	0.39	-0.11	0.04	-0.32	-0.02
TEMP_RANGE	0.11	-0.48	0.03	0.53	-0.13	-0.45	-0.2	0.04
ISOTHERM	-0.14	0.13	-0.19	-0.48	0.1	-0.39	0.13	0.03
TEMP_SEAS	0.17	-0.3	0.19	0.54	-0.1	0.24	-0.12	0.07
MAX_T	-0.13	-0.46	-0.15	0.49	-0.16	-0.03	-0.3	-0.01
MIN_T	-0.23	-0.32	-0.33	-0.54	0.06	0.26	0	-0.05
ANN_T_RANGE	0.13	-0.49	0.12	0.53	-0.12	-0.14	-0.18	0.03
MEAN_T_WtQ	-0.05	-0.25	-0.11	0.5	-0.19	0.17	-0.31	0.19
MEAN_T_DQ	-0.32	-0.4	-0.35	-0.53	0.05	0.12	-0.05	-0.09
MEAN_T_WaQ	-0.13	-0.41	-0.11	0.5	-0.14	0.12	-0.32	0.01
MEAN_T_CQ	-0.3	-0.42	-0.34	-0.51	0.01	0	-0.09	-0.24
ANN_PRC	-0.22	0.29	-0.11	-0.53	0.15	0.32	0.14	-0.05
PRC_WtM	-0.26	0.12	-0.19	-0.46	0.11	0.05	0.11	-0.29
PRC_DM	-0.21	0.4	-0.27	-0.48	0.09	0.22	0.13	0.06
PRC_SEAS	0.11	-0.4	0.07	0.48	-0.11	-0.33	-0.15	-0.03
PRC_WtQ	-0.23	0.11	-0.12	-0.37	0.13	0.17	0.13	-0.28
PRC_DQ	-0.19	0.49	-0.27	-0.46	0.08	0.21	0.12	0.08
PRC_WaQ	-0.2	0.5	-0.41	-0.51	-0.08	0.22	-0.04	0.04
PRC_CQ	-0.09	0.05	-0.27	-0.54	0	0.05	0.13	-0.14

Chapitre III : Variabilité des constituants chimiques des graines

Les analyses de régressions linéaires au sein des régions Ouest et Est entre les différentes formes de tocophérols et les variables climatiques, géographiques et topographiques sont résumé dans les **Tableaux III.27** et **III.28**. Les résultats obtenus montrent des corrélations faibles et peu significatives pour les teneurs en α -tocophérol et en γ -tocophérol avec des coefficients de corrélation de Pearson r inférieurs à 0,5 et supérieurs à -0,5. Les corrélations obtenues entre les variables climatiques et les deux formes minoritaires de tocophérols sont très peu significatives au sein des deux grandes régions. On observe au sein du transect Ghana-Burkina des corrélations plus importantes qu'au sein du transect Mali, bien que celles-ci soient très faibles avec des coefficients inférieurs à 0.6.

Tableau III.28. Graphiques de corrélations entre les deux principaux tocophérols et les variables climatiques et topographiques.





4 DISCUSSION

4.1 Composition de la matière grasse du Karité par rapport à la littérature

Dans cette étude nous avons confirmé les très fortes teneurs en matière grasse, représentant entre 30 et 60% de la matière sèche totale des graines du Karité, le reste de la graine se composant des constituants glucidiques et protéiques (**Figure III.8**). Ces teneurs en matière grasse sont comparables notamment à celles mesurées dans les fèves de cacao (*Theobroma cacao* L.) autour de 50% de la matière sèche (Liendo *et al.*, 1997).

Les 17 sites étudiés dans ce travail peuvent être considérés comme un échantillon représentatif de la variabilité sur l'aire car il couvre globalement la gamme de variation observée dans le projet ProKarité (Maranz & Wiesman, 2003; Maranz *et al.*, 2003; Maranz *et al.*, 2004; Di Vincenzo *et al.*, 2005) notamment pour les teneurs en matière grasse, et les proportions relatives dans les différents en acides gras. La **Figure III.12** en particulier illustre la comparaison entre les proportions relatives en acide oléique et en acide stéarique obtenues dans l'étude ProKarité (Maranz & Wiesman, 2003; Maranz *et al.*, 2003; Maranz *et al.*, 2004; Di Vincenzo *et al.*, 2005) et celles obtenues dans ce travail de thèse (projet européen InnovKar). Cette étude confirme ainsi la prédominance de l'acide stéarique (saturé ; C18 :0) entre 26 et 48% et de l'acide oléique (mono-insaturé ; C18 :1 n-9) entre 38 et 60%. Par ailleurs, on constate que la somme de ces deux acides gras constitue invariablement entre 85 et 90% de la matière grasse. Ce résultat s'explique de par la nature même de la voie de biosynthèse de ces composés (**Figure III.1**). Ainsi, ces deux composés à chaînes de 18 carbones sont interdépendants, car l'acide oléique est directement issu de la réaction de désaturation de l'acide stéarique, par une enzyme soluble du stroma plastidial, nommée Δ^9 -stéaroïl-acyl-carrier-protéine désaturase (SAD) (Lindqvist *et al.*, 1996; Somerville *et al.*, 2000). Ceci explique donc la corrélation forte entre ces deux acides gras majoritaires. Si les coefficients de variation observés pour les deux acides gras majoritaires sont relativement faibles (14% pour l'acide stéarique ; et 10% pour l'acide oléique).

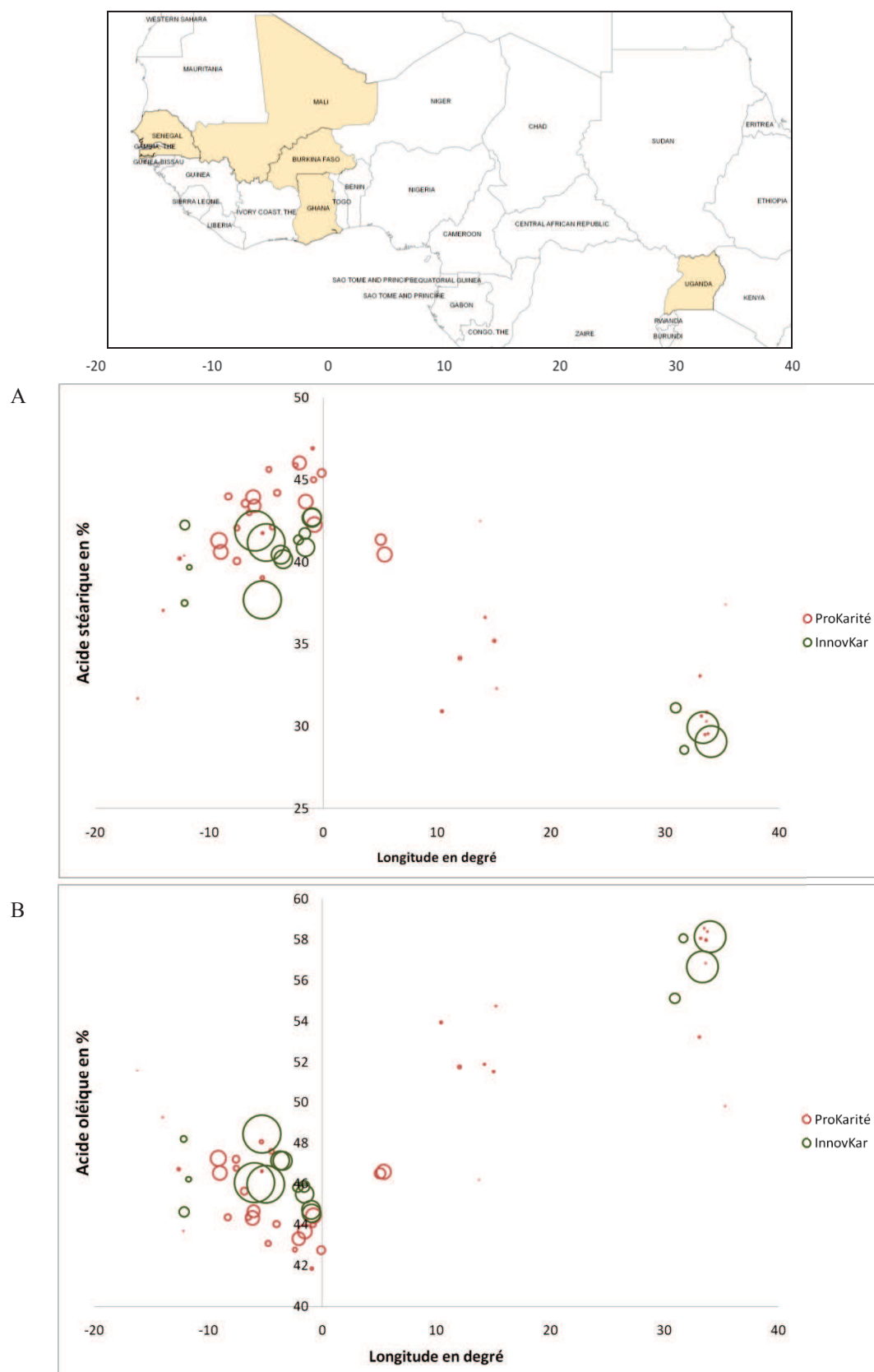


Figure III.12. : Comparaison des profils d'acides gras obtenu sur l'aire naturelle dans le projet ProKarité (Maranz & Wiesman, 2003; Maranz *et al.*, 2004) et dans de cette étude (projet InnovKar) pour : A) proportions relatives en acide Stéarique par site; B) proportions relatives en acide Oléique par site. Le diamètre des bulles est proportionnel au nombre d'arbre échantillonné au sein du site.



Chapitre III : Variabilité des constituants chimiques des graines

Les variations observées induisent cependant de grand bouleversement sur le plan physico-chimique. En effet, la température de fusion de l'acide stéarique est de 69,6 °C tandis que la température de fusion de l'acide oléique n'est que de 16,3 °C (Bailey, 1982). La part relative de ces deux composés dans la matière grasse influence donc fortement la consistance de l'huile extraite. Ainsi, plus les huiles de karité sont riches en acide stéarique plus elles sont solides à température ambiante et inversement pour l'acide oléique. Ainsi, l'Afrique de l'Ouest commercialise un beurre de karité à l'état solide alors qu'à l'Est de l'Aire naturelle de l'espèce, la matière grasse extraite est liquide et vendue en bouteille. De même que pour les teneurs en matière grasse, la composition en acides gras est comparable à celle du beurre de cacao (Liendo *et al.*, 1997), justifiant notamment le classement du beurre de Karité en tant que matière grasse végétale de substitution au beurre de Cacao (CBO – *Cocoa Butter Equivalent*).

Concernant la partie insaponifiable analysée que constituent les tocophérols, les valeurs obtenues ne confirment que partiellement les résultats publiés antérieurement (**Figure III.13**). Ainsi, les teneurs en γ -tocophérols ainsi que les β - et δ -tocophérols mesurées pour le projet ProKarité (**Figure III.13**), sont en moyenne 2 à 3 fois plus importantes (Maranz & Wiesman, 2004) que celles observées dans notre étude. Cette différence peut s'expliquer par le nombre d'individus mesurés beaucoup plus important dans notre étude (en moyenne 3,5 fois plus d'individus par site) et un échantillonnage plus équilibré entre les sites et statistiquement plus fiable (**Figure III.13**). En particulier, sur les 20 sites récoltés dans le projet ProKarité, 12 présentent un échantillonnage très limité de seulement 1 ou 2 individus. Une deuxième explication possible de cette différence, est la faible répétabilité de la mesure des teneurs en tocophérols. En effet, ces composés liposolubles qui ne représentent que moins de 1% de la matière grasse totale montrent dans la littérature des incertitudes de mesures importantes. En particulier, on observe chez le noyer (*Nux regia* L.) des écarts de mesures de 5 à 10% (Lavedrine *et al.*, 1997). Dans notre étude nous avons estimé ces erreurs de mesure en laboratoire à environ 9% pour l' α -tocophérol, 13% pour le β -tocophérol, 10% pour le γ -tocophérol et plus de 18% pour le δ -tocophérol. Dans notre étude, chaque échantillon a donc été analysé trois fois afin de limiter l'impact de l'incertitude sur la mesure retenue. Or dans les études précédentes une seule mesure a été effectuée par échantillon, limitant ainsi la robustesse statistique des mesures effectuées.

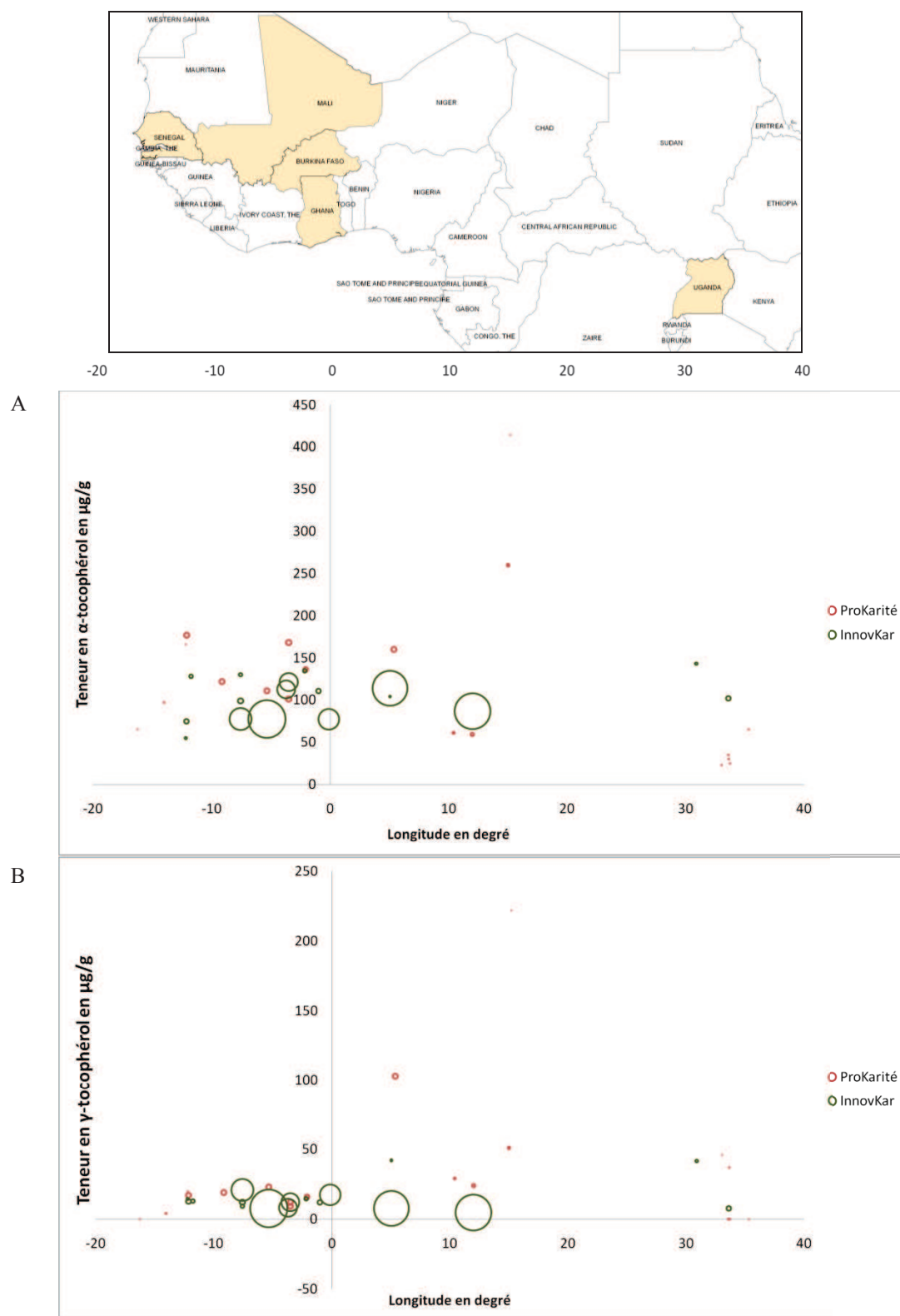
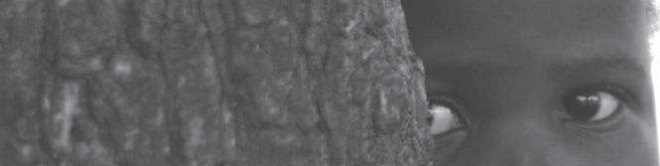


Figure III.13 : Comparaison des teneurs en tocophérols mesurées sur l'aire naturelle dans le projet ProKarité (Maranz & Wiesman, 2003; Maranz *et al.*, 2004) et dans de cette étude (projet InnovKar) pour : A) teneur en α -tocophérol par site; B) teneur en γ -tocophérol par site. Le diamètre des bulles est proportionnel au nombre d'arbre échantillonné au sein du site.



Enfin, une dernière explication possible vient de la fonction même de ces composés. Ainsi, ces molécules antioxydantes (Brigelius-Flohé, 2009) sont soumis à des dégradations importantes, notamment du fait des dérivés réactifs de l'oxygène (DRO) (*Reactive oxygen species, ROS*) comme le peroxyde d'hydrogène (H_2O_2). Il n'est donc pas exclu que les teneurs faibles mesurées dans notre étude soit dues à des dégradations post-récolte des tocophérols. Afin de vérifier cela, l'indice de peroxydation des graines est en cours d'analyse et les résultats seront analysés début 2011.

Cependant, cette étude révèle que les teneurs moyennes en tocophérols sont similaires à celles d'autres espèces des tropiques (Ching & Mohamed, 2001) comme *Apium graveolens* un celeri local (α -tocophérol = 136,4 $\mu\text{g/g}$), ou encore le thé noir (*Camelia chinensis* ; α -tocophérol = 183,3 $\mu\text{g/g}$), mais inférieures aux teneurs mesurées chez le palmier à huile (*Elaeis guineensis* ; α -tocophérol = 300 à 500 $\mu\text{g/g}$) ou le tournesol (*Helianthus annuus* ; α -tocophérol = 500 à 700 $\mu\text{g/g}$) (Grusak & Dellapenna, 1999). Ces teneurs importantes notamment en α - et γ -tocophérols participent ainsi à ralentir les phénomènes de peroxydation lipidique qui peuvent altérer à terme les biomolécules et mettre en péril le pouvoir germinatif de la graine (Brigelius-Flohé, 2009). Ces composés confèrent donc au beurre de Karité un potentiel intéressant pour l'industrie cosmétique, qui exploite ses propriétés anti-oxydantes notamment pour l'élaboration de crème combattant les effets des formes réactives de l'oxygène, telles que les soins « anti-âges » (Collinson & Zedie-Bosuener, 1999).

La variation des constituants chimiques de la graine de Karité sont donc importants que ce soit pour les profils d'acides gras ou encore pour la fraction insaponifiable des lipides que nous avons étudié. On peut cependant se demander quelle part de cette variation est liée à des effets génétiques et quelle part à des effets environnementaux.

4.2 Patron de variation interrégionale des composants lipidiques de la graine de Karité

Pour l'ensemble des composés analysés, nous avons mis en évidence une structuration régionale (Est-Ouest) extrêmement forte expliquant jusqu'à 87% de la variation chimique des graines, notamment en acide oléique, en acide stéarique et en γ -



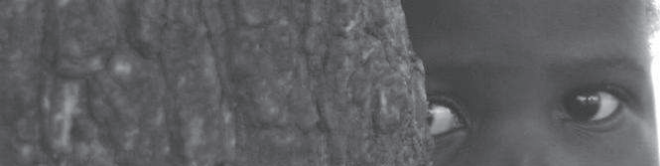
Chapitre III : Variabilité des constituants chimiques des graines

tocophérol. La composition lipidique des graines distingue donc clairement les provenances d'Afrique occidentale et d'Afrique de l'Est. Plusieurs facteurs peuvent expliquer ce patron.

En outre, cette différence marquée entre Afrique de l'Est et Afrique de l'Ouest est cohérente avec la distinction sur l'aire naturelle des deux sous-espèces *nilotica* à l'Est et *paradoxa* à l'Ouest (Hall *et al.*, 1996). Si cette distinction en sous-espèces se base sur des considérations discutables, nous mettons en évidence dans le Chapitre 2 de cette thèse qu'une rupture ancienne dans l'aire naturelle de l'espèce durant le dernier maximum glaciaire (20000 ans avant notre ère) pourrait être à l'origine d'une forte structuration génétique séparant l'aire de distribution en deux au niveau des monts Adamawa à la frontière entre le Cameroun et le Nigéria. Si cette hypothèse n'est pas vérifiable dans notre étude car notre échantillonnage ne contient pas de populations intermédiaires entre l'Ouganda et l'Afrique de l'Ouest, les résultats apportés par le projet ProKarité soutiennent cette théorie. Ainsi, bien que souffrant d'un échantillonnage réduit, on observe pour les profils d'acides gras que les populations du Cameroun, à l'Est des monts Adamawa (longitude $\geq 10^\circ$ Est) semblent être de type « Ouganda », alors que les populations se situant juste à l'Ouest de ces montagnes au Nigéria (longitude $\leq 10^\circ$ Est) montrent des profils d'acides gras plutôt types « Afrique de l'Ouest » (**Figure III.12**).

Ce patron de variation peut également résulter des différences entre les conditions environnementales. Si au sein des régions, il semble que les régimes climatiques soient peu variables, les provenances d'Afrique de l'Est et d'Afrique de l'Ouest sont caractérisées par des régimes climatiques distincts, notamment pluviométriques (**Tableaux III.3** et **III.4**). Ainsi la structure chimique observée entre ces deux régions pourrait résulter d'effets purement environnementaux ou d'une adaptation région-spécifique des populations.

L'échantillonnage en milieu naturel non-contrôlé ne nous permet cependant pas de vérifier ces hypothèses. Notamment, plusieurs facteurs évolutifs sont confondus à cette échelle. La dérive génétique notamment entre ces deux régions est forte (Chapitre 2) et renforcée par un éloignement géographique important. Aucune conclusion ne peut donc être émise concernant l'origine des effets génétiques expliquant les différences Est-Ouest.



4.3 Patrons de variation intrarégionaux des composants lipidiques de la graine de Karité

Dans cette étude, nous avons mis en évidence que la part de la variance intersites au sein des régions est très dépendante du caractère chimique observé et de la région considérée. Cependant, nous avons observé pour les caractères liés aux acides gras et l' α -tocophérol, des variances intersites plus importantes au sein de la région Ouest (entre 10 et 57%) qu'au sein de la région Est (entre 0 et 39%). Ce résultat peut s'expliquer par le fait qu'un échantillonnage plus étendu et plus conséquent a été effectué à l'Ouest de l'aire naturelle par rapport à l'Est limité à l'Ouganda. Toutefois, concernant les β -, γ - et δ -tocophérols, les variances intersites se sont révélées très faibles au sein de l'Afrique de l'Ouest (entre 5 et 8%) alors qu'elles sont très fortes (entre 47 et 68%) à l'Ouest. Ce résultat suggère un impact potentiellement fort de ces tocophérols dans la stratégie adaptative au sein de la zone Est. Cependant, ce résultat est à relativiser car les coefficients de variation sont très forts pour ces composés et l'échantillonnage au sein de la zone Est et le nombre d'individus caractérisé restent limités (entre 4 et 9 par site).

La variance intersites observée pour les composants majeurs de la matière grasse en Afrique occidentale est similaire à celles observées pour certains composants chimiques des huiles essentielles de Santal (*Santalum austrocaledonicum*) (Bottin *et al.*, 2007) avec par exemple une variance intersites de 23% pour l' α -Santalol ou encore 35% pour le Nuciferol. Elle est aussi semblable à celles observées sur d'autres caractères quantitatifs chez *Eucalyptus urophylla*, notamment la taille (43%), la circonférence (34%) ou encore le taux de mortalité (14%) (Tripiiana *et al.*, 2007).

La littérature montre que la variabilité de constituants chimiques des graines, notamment pour les acides gras et les tocophérols, peut être expliquée génétiquement pour de nombreuses espèces, et en particulier pour les espèces de grandes cultures. La composition en acides gras des graines notamment est soumise à une forte dépendance génétique avec des héritabilités jusqu'à 88% chez le soja (Rebetzke *et al.*, 1998; Primomo *et al.*, 2002; Hou *et al.*, 2006), la cacahuète (Amaral *et al.*, 2004) ou encore le noyer (Isleib *et al.*, 2006).



Si notre plan d'expérience (échantillonnage en milieu naturel) ne nous permet pas de calculer l'héritabilité des caractères étudiés, il a été montré chez le Karité (Sanou *et al.*, 2005) une forte héritabilité pour certains caractères agro-morphologiques, liés notamment à la croissance juvénile, résultant de mécanismes d'adaptation locale. Ainsi, bien que l'hypothèse ne puisse être vérifiée pour les constituants chimiques des graines de Karité, le patron de variation intersite observée en Afrique de l'Ouest pour certains caractères pourrait être lié à l'adaptation locale.

4.4 Relation avec les variables environnementales

La morphologie et la composition chimique des propagules des plantes est un caractère supposé fortement soumis aux pressions de sélection et à l'adaptation à l'environnement (Linder, 2000). En particulier, la composition de la graine en acides gras est prépondérante dans les mécanismes de tolérance à la sécheresse et à de fortes ou faibles températures, notamment dans le maintien de l'intégrité des parois cellulaires (Linder, 2000; Khurana & Singh, 2001; Casini *et al.*, 2003; Pritchard *et al.*, 2004). Les tocophérols et notamment l' α -tocophérol, sont eux des facteurs majeurs du système de défense antioxydant des plantes (Brigelius-Flohé, 2009). Ces composés membranaires lipophiles inhibent l'oxydation d'autres molécules (tels que les acides gras notamment), stoppant la progression des chaînes de réaction oxydative en captant l'électron libre des formes réactives de l'oxygène (ROS) (Brigelius-Flohé, 2009). Or, ces ROS apparaissent avec la survenue de stress tels que le déficit hydrique, les fortes et basses températures, ou encore un excès de luminosité. Ainsi, le système de défense antioxydant constitue un caractère adaptatif important pour la protection des plantes, des graines encore non germées et des jeunes plantules (Vertucci & Leopold, 1987; Kamal-Eldin & Appelqvist, 1996b; Velioglu *et al.*, 1998; Klein & Kurilich, 2000).

L'hypothèse apportée par Marranz est que la composition chimique des graines de karité en acides gras (Maranz & Wiesman, 2003) et en tocophérols (Maranz & Wiesman, 2004) suit des variations environnementales. En particulier, il propose que la quantité d'acide oléique augmente avec la pluviométrie (inversement pour l'acide stéarique) (Maranz & Wiesman, 2003; Maranz *et al.*, 2004), et que les teneurs en α -tocophérol augmentent avec la température du site d'origine. Dans notre étude nous avons donc tenté de vérifier



l'hypothèse suggérée par Maranz *et al.* (2003, 2004) de relation entre variations environnementales et composition des graines de Karité, en utilisant un échantillonnage conséquent, en analysant séparément les deux sous-espèces (les deux régions : Afrique de l'Ouest et Afrique de l'Est) et en échantillonnant des populations réparties sur deux clines environnementaux (transects Mali et Ghana-Burkina) au sein de la zone Ouest.

Au sein des deux grandes zones échantillonnées, nous avons montré que les relations entre variation chimique des graines de Karité et variation climatique des sites de récolte sont très faibles et l'hypothèse émise par Maranz *et al.* (2003, 2004) n'a pas non plus été vérifiée au sein des deux transects étudiés, que ce soit pour les tocophérols ou les compositions en acides gras. On note cependant que les variables environnementales utilisées pour décrire nos populations montrent de faibles variations (**Tableau III.3 et III.4**). Il n'est donc pas étonnant que les modèles de régression utilisés n'aient pas permis de mettre en évidence de corrélations fortes.

Notre étude montre que la fertilité du sol semblerait avoir un impact positif dans le remplissage des graines en matière grasse au sein des deux grandes régions Est et Ouest. Il est cependant possible que cet effet soit confondu avec d'autres facteurs biotiques ou abiotiques.

4.5 Variations morphologiques et relation avec les constituants chimiques des graines au Mali

Dans cette thèse, nous nous sommes également intéressés à la variation morphologique des semenciers ainsi que des graines échantillonnées sur le transect Mali. Les caractères morphologiques liés à la taille des arbres ont montré des coefficients de variation (entre 19 et 35%) proches de ceux obtenus dans d'autres études (Sanou *et al.*, 2006) ou encore pour d'autres espèces forestières pour des caractères similaires (Tripiana *et al.*, 2007). L'analyse de la morphologie de ces arbres au Mali montre une relation Nord-Sud. Ce résultat s'explique en particulier un couvert végétal moins dense au Nord du fait d'un climat plus xérique. De plus en allant vers le Nord, l'histoire de l'exploitation des karités par l'Homme en parc agroforestiers serait plus ancienne et la pression démographique humaine serait plus forte, réduisant de plus en plus les cycles de jachère au besoin de la production agricole (Lovett & Haq, 2000a).



Chapitre III : Variabilité des constituants chimiques des graines

Dans notre étude, nous avons montré que les coefficients de variation associés aux caractères morphologiques des fruits, et notamment au poids des fruits (variable : PTF ; PNF ; PNS) sont similaires à ceux observés pour les caractères liés à la taille des arbres (entre 24 et 35%). Nous avons mis en évidence que ces caractères ne sont que très faiblement corrélés à la composition chimique des graines.

Si les résultats obtenus ne permettent pas de conclure, on peut penser que les populations étudiées résultent plus d'adaptations locales que clinales, et que ces mécanismes ne s'expriment pas de façon pléiotropique.

4.6 Impact de la domestication sur la variabilité chimique

Dans ce chapitre, nous avons mis en évidence une plus grande variabilité des constituants chimiques sur le transect « Mali » que sur le transect « Ghana-Burkina ». Cette différence est peut-être liée à un échantillonnage un peu plus important sur le transect Mali. Il est cependant probable que cette différence soit liée à l'histoire de la domestication de l'espèce. Ainsi, le transect Ghana-Burkina se situe principalement sur le plateau de Mossi à cheval sur le Burkina et le Ghana (**Figure III.5**). Notamment, des études palynologiques ont mis en évidence des preuves d'activités agropastorales dans cette région il y a 1000 ans (Ballouche & Neumann, 1995). Plus encore, des résidus de coques de noix et de charbon de Karité ont été découverts dans un village mis à jour, datant de 1000 ans (Neumann *et al.*, 1998). Cette hypothèse de domestication et de sélection anthropique du Karité dans cette région a également été vérifiée dans divers études morphologiques de l'espèce (Lovett & Haq, 2000a; Maranz & Wiesman, 2003) et pourrait ainsi expliquer la faible diversité observée dans cette région en comparaison du Mali voisin.

5 CONCLUSION & PERSPECTIVES

Dans ce chapitre nous avons mis en évidence la forte structuration de la variabilité chimique à l'échelle de l'aire naturelle. Nous suggérons que cette structure, cohérente avec l'histoire évolutive du Karité décrite dans le chapitre 2, résulterait de mécanismes adaptatifs et de la dérive génétique. Au sein de la région Ouest de l'aire naturelle, nous avons mis en évidence une structure contrastée de la variation chimique en fonction des caractères, suggérant une adaptation locale des populations. Le plan d'expérience mis en place, ne nous



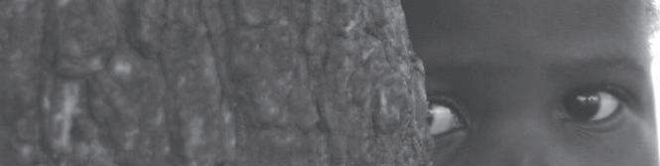
Chapitre III : Variabilité des constituants chimiques des graines

a cependant pas permis pas de conclure et vérifier ces hypothèses. En particulier, nous avons été confrontés au problème du calcul de l'héritabilité en milieu naturel sans connaître le pedigree. Ainsi, ne pouvant dissocier la part des effets environnementaux, génétiques et de l'interaction génotype X environnement, nous proposons ici des pistes possibles afin de mieux comprendre la variation chimique des constituants chimiques de la graine de Karité et mesurer l'origine génétique de cette variation.

Ainsi, nous proposons de densifier l'échantillonnage en nombre de sites mais également en nombre d'individus analysés, ce notamment pour les tocophérols souffrant d'un échantillonnage plus limité. Ceci permettrait d'avoir une vision plus nette de la variation des constituants, et ce en particulier pour les zones non échantillonnées en Afrique Centrale et en Afrique de l'Ouest. Ceci permettrait de plus de vérifier l'impact de la structuration génétique de l'espèce liée aux bouleversements climatiques de la période Quaternaire mise en évidence dans le Chapitre 2, dans l'expression de la variation des caractères chimiques sur l'aire naturelle. Nous proposons également au sein des régions de redéfinir des transects plus pertinents afin notamment de couvrir des plages de variation climatique plus marquées.

Cependant, la difficulté de mettre en place les missions de récoltes sur place (en écosystème naturel) et le coût important en temps et en argent des analyses chimiques en laboratoire constituent les principaux obstacles à l'augmentation de l'échantillonnage. Afin de limiter les coûts d'une augmentation du nombre d'échantillon à analyser, nous avons au cours de ce travail de thèse, développé un outil de prédiction de la composition des graines de Karité fiable et peu coûteux (Davrieux *et al.*, 2010). Voir section « Valorisation ».

D'autre part, des outils d'estimation de l'héritabilité en milieu naturel existent et pourront être utilisés (Ritland & Travis, 2004). Nous avons testé ce type de méthodologie avec l'estimation de l'héritabilité par l'approche de Ritland au sein d'une parcelle test à Mperesso au Mali (Bouvet *et al.*, 2008). Cependant, cette étude a révélé que ces méthodes basées sur des modèles de régression ne sont pas bien adaptées aux populations d'arbres en milieux agroforestiers. Il serait cependant intéressant avec les nouvelles méthodes haut débit de reprendre ce type d'expérimentation en utilisant des marquages conséquents (Bessega *et al.*, 2009).



Chapitre III : Variabilité des constituants chimiques des graines

Enfin, la solution la plus intéressante mais également la plus coûteuse et laborieuse serait la mise en place d'un dispositif d'essai d'interaction génotype x environnement. Ainsi, ce test de provenances en milieu contrôlé permettrait d'évaluer l'impact respectif des différents effets pouvant expliquer la variation des constituants chimiques des graines de karité. Cependant, comme nous l'avons décrit dans le chapitre I, le Karité est une espèce à croissance lente, or ce type de dispositif nécessite d'attendre que les individus soient arrivés à maturité reproductive, c'est-à-dire à l'âge de 15 à 20 ans.

Enfin, les acides gras que nous avons étudiés au cours de ce travail sont rarement présents sous forme libre. La quasi-totalité de ces composés s'organise au sein des graines dans des molécules plus grosses : les triglycérides (TAG). Ainsi, étudier ces composés permettrait d'avoir une vision plus claire des variations sur le plan de la physiologie des arbres. Ces constituants chimiques sont en cours d'analyse et l'étude de la variation des TAG sera effectuée en 2011.

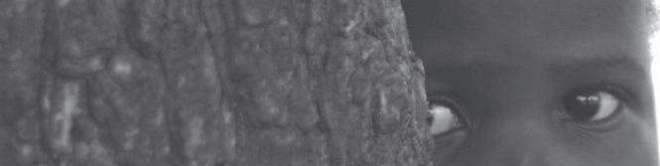


6 PRÉDICTION SPIR DE LA COMPOSITION CHIMIQUE DES FRUITS DU KARITÉ

Near Infrared Spectroscopy for High-Throughput Characterization of Shea Tree (*Vitellaria paradoxa*) Nut Fat Profiles

Objectif. Le karité (*Vitellaria paradoxa*) est un arbre majeur des systèmes agroforestiers d'Afrique Soudano-Sahélienne depuis plus de 1000 ans (Lovett & Haq, 2000b). Son exploitation aujourd'hui croissante avec plus de 600000 tonnes de noix sèches par an, et la production de beurre de Karité à partir de ces noix a très fortement augmenté avec la mise en place d'un commerce international. Ce beurre de Karité offre ainsi une opportunité de développement durable intéressante pour dans les pays de la zone soudanienne et un potentiel attractif pour les industries alimentaire et cosmétique (Collinson & Zedie-Bosuener, 1999). Afin de répondre à cette demande croissante, la filière Karité nécessite aujourd'hui le développement d'outils spécifiques de notamment de contrôle qualité des graines commercialisées pour la fabrication du beurre. Le but de cette étude est donc le développement d'un outil de prédiction de la composition des noix de karité par spectroscopie proche-infrarouge (SPIR ou en anglais : *NIRS = Near-InfraRed Spectroscopy*) afin de caractériser rapidement et à bas coût les noix de karité (**Annexe III.1**). La technique SPIR est basée sur les propriétés vibrationnelles de molécules sous l'effet de différentes longueurs d'onde infrarouge. Les spectres d'absorption sont ensuite corrélés spécifiquement avec les différentes teneurs en composés biochimiques (Pasquini, 2003).

Matériel et Méthodes. Les noix récoltées sur 624 arbres dans cinq pays africains (Sénégal, Mali, Burkina Faso, Ghana et Ouganda), ont été caractérisés chimiquement pour la teneur en matière grasse, la teneur en eau et les profils d'acides gras par chromatographie en phase gazeuse (CPG). Les spectres d'émission proche-infrarouges (entre 400 et 2500 nm de longueur d'onde λ) des noix réduites en poudre de 0,5 à 0,8 mm de chaque échantillon ont ensuite été mesurés en utilisant un NIRS 6500 monochromator (Foss NIRSystems, Silver Spring, MD). Une analyse en composantes principales (ACP) a été effectuée afin de décrire la diversité des profils spectraux des échantillons analysés. Dix pour cent des échantillons (60) ont été choisis au hasard pour la validation des modèles et les échantillons restants ont été utilisés pour la calibration des équations de prédictions. Pour chaque constituant mesuré (teneur en eau, teneur en matière grasse, proportions relatives en acide palmitique,

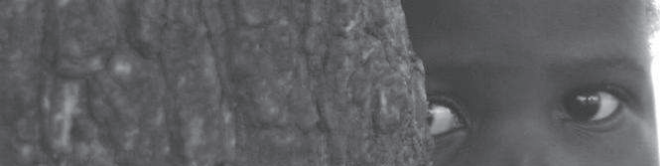


Chapitre III : Variabilité des constituants chimiques des graines

stéarique, oléique, vaccénique, linoléique, linolénique et arachidique) une équation de calibration a été développée à l'aide d'un modèle modifié de régression des moindres carrés : *Modified Partial Least Square* (MPLS). Les performances des modèles développés ont été évaluées en utilisant le *Ratio Performance to Deviation* (RPD_p) et le coefficient de détermination R_p^2 , paramètres obtenus par comparaison sur l'échantillon de validation, des valeurs prédites par les modèles et des valeurs mesurées en laboratoire.

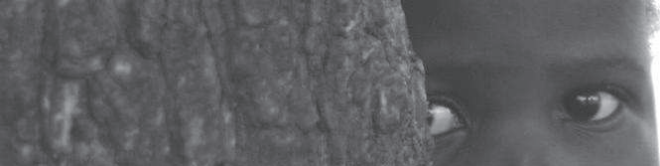
Résultats & Discussion. L'étude préliminaire de la variabilité spectrale des poudres de graines de Karité par analyse en composantes principales montre une discrimination très forte des provenances d'Afrique de l'Ouest de celles d'Afrique de l'Est. Les calibrations obtenues pour les teneurs en eau ($RPD_p = 4,45$; $R_p^2 = 0,95$) et les teneurs en matière grasse ($RPD_p = 5,6$; $R_p^2 = 0,97$) sont très bonnes et permettent de prédire ces caractéristiques de façon précise. Les modèles de prédiction proche-infrarouge développés pour les proportions relatives en acide stéarique ($RPD_p = 6,26$; $R_p^2 = 0,98$) et acide oléique ($RPD_p = 7,91$; $R_p^2 = 0,99$) dans la matière grasse des graines de karité sont extrêmement performant et permettent une prédiction de ces deux acides gras majoritaires (~90% des acides gras totaux) du beurre de karité avec une précision similaire à celle des méthodes de chromatographie en phase gazeuse. Enfin, les équations de calibration obtenue pour la prédiction des acides gras minoritaires de la matière grasse sont beaucoup moins performantes : acide palmitique ($RPD_p = 1,3$; $R_p^2 = 0,45$), acide vaccénique ($RPD_p = 1,19$; $R_p^2 = 0,23$), acide linoléique ($RPD_p = 1,71$; $R_p^2 = 0,68$), acide linolénique ($RPD_p = 1,3$; $R_p^2 = 0,39$) et acide arachidique ($RPD_p = 1,5$; $R_p^2 = 0,57$). Cependant les résultats obtenus pour les acides linoléique et palmitique permettent une estimation grossière de ces deux acides gras représentant respectivement 7% et 1% de la quantité totale d'acides gras. Enfin, nous avons obtenu une équation de prédiction très précise pour le ratio entre acides gras saturés et acides gras insaturés ($RPD_p = 5,72$; $R_p^2 = 0,97$).

Principales conclusions. Cette étude a démontré la capacité de la spectroscopie proche infrarouge (SPIR) pour la détermination des proportions relatives des deux acides gras majoritaires, de la teneur en eau et de la teneur en matière grasse des graines de Karité. Le coût très faible ainsi que la vitesse d'analyse (100 échantillons par jour contre une dizaine en chimie classique) permet ainsi d'envisager un phénotypage à haut débit des noix de karité. Les modèles ainsi développés répondent aux attentes des industriels notamment en



Chapitre III : Variabilité des constituants chimiques des graines

permettant un calcul très rapide des ratios d'acides gras saturés et insaturés, indispensable afin de prévoir la fluidité et les propriétés physico-chimiques des beurres extraits. Enfin cette technique « haut débit » constitue un outil très performant afin d'étudier les patrons de variation de ces caractères au sein de l'espèce en relation notamment avec des gradients environnementaux et en association avec des variations nucléotidiques de gènes impliqués dans les voies de biosynthèses de ces composés.



Chapitre III : Variabilité des constituants chimiques des graines

Near Infrared Spectroscopy for High-Throughput Characterization of Shea Tree (*Vitellaria paradoxa*) Nut Fat Profiles

FABRICE DAVRIEUX,^{*,†} FRANÇOIS ALLAL,[‡] GEORGES PIOMBO,[§] BOKARY KELLY,[~]
JOHN B. OKULO,[^] MASSAMBA THIAM,[#] OUSMANE B. DIALLO,[†] AND JEAN-MARC BOUVET[‡]

[†]CIRAD, Performance of Tropical Production and Processing Systems Department, UMR QUALISUD, TA B-95/16, 73, avenue Jean François Breton, 34398 Montpellier Cedex 5, France, [‡]CIRAD, Biological Systems Department, Research Unit 39, "Genetic Diversity and Breeding of Forest Tree Species", Campus international de Baillarguet TA A-39/C, 34398 Montpellier Cedex 5, France, [§]CIRAD, Performance of Tropical Production and Processing Systems Department, UMR IATE, TA B 62/16, 73 rue Jean François Breton, 34398 Montpellier Cedex 5, France, [~]Institut d'Economie Rurale, PRF/CRRA-SIKASSO BP 178/16, Sikasso, Mali, [^]Department of Forest Biology & Ecosystems Management, Faculty of Forestry and Nature Conservation, Makerere University, PO Box 7062 Kampala, Uganda, [#]Institut Sénégalais de la Recherche Agronomique, Centre National de la Recherche Forestière, BP 2312 Dakar-Hann, Senegal, and [†]INERA-Centre National de la Recherche Scientifique et Technologique, 03 BP 7041, Ouagadougou, Burkina Faso

The Shea tree (*Vitellaria paradoxa*) is a major tree species in African agroforestry systems. Butter extracted from its nuts offers an opportunity for sustainable development in Sudanian countries and an attractive potential for the food and cosmetics industries. The purpose of this study was to develop near-infrared spectroscopy (NIRS) calibrations to characterize Shea nut fat profiles. Powders prepared from nuts collected from 624 trees in five African countries (Senegal, Mali, Burkina Faso, Ghana and Uganda) were analyzed for moisture content, fat content using solvent extraction, and fatty acid profiles using gas chromatography. Results confirmed the differences between East and West African Shea nut fat composition: eastern nuts had significantly higher fat and oleic acid contents. Near infrared reflectance spectra were recorded for each sample. Ten percent of the samples were randomly selected for validation and the remaining samples used for calibration. For each constituent, calibration equations were developed using modified partial least squares (MPLS) regression. The equation performances were evaluated using the ratio performance to deviation (RPD_p) and R_p² parameters, obtained by comparison of the validation set NIR predictions and corresponding laboratory values. Moisture (RPD_p = 4.45; R_p² = 0.95) and fat (RPD_p = 5.6; R_p² = 0.97) calibrations enabled accurate determination of these traits. NIR models for stearic (RPD_p = 6.26; R_p² = 0.98) and oleic (RPD_p = 7.91; R_p² = 0.99) acids were highly efficient and enabled sharp characterization of these two major Shea butter fatty acids. This study demonstrated the ability of near-infrared spectroscopy for high-throughput phenotyping of Shea nuts.

KEYWORDS: *Vitellaria paradoxa*; NIRS; fatty acids; high-throughput characterization; shea butter; MPLS regression

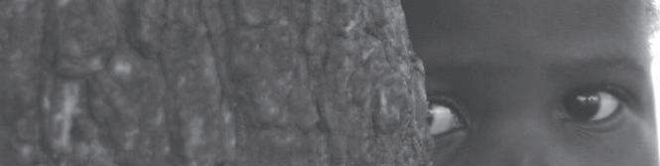
INTRODUCTION

The Shea tree (*Vitellaria paradoxa* C. F. Gaertn, Sapotaceae family), called karité in French, is a major tree species of African agroforestry systems. The native range of this long-lived (over 200 years) savannah tree species is a large belt 6000 km long from East (Senegal) to West (Uganda) and 600 km wide from north of the equator to south of the Sahara. *Vitellaria paradoxa* is divided into 2 subspecies: subsp. *paradoxa* is distributed throughout most of the range (including West and Central Africa) while subsp. *nilotica* is found in the eastern part of Sub-Saharan Africa. Within

that wide distribution, there is an immense ecological diversity, with altitudes ranging from almost 0 to nearly 1500 m, a mean annual temperature of between 22 and 30 °C and a mean annual rainfall of 400 mm to 1400 mm. Shea trees reach a maximum height of 20 m with a trunk diameter ranging generally from 0.3 to 1 m at breast height. Annual flowering and fruiting occurs from December to May with local variations (1).

The Shea tree has been exploited by African communities for about 3000 years and offers an opportunity for sustainable development in Sudanian countries (2, 3). Indeed, Shea butter extracted from fruit kernels provides an attractive potential for both the food and cosmetics industries (1). The estimated yield of dry kernel is about 600,000 tons per year, and exports have

* Corresponding author. Tel: þ 33 467 615 893. Fax: þ 33 467 614 433. E-mail: davrieux@cirad.fr.



Chapitre III : Variabilité des constituants chimiques des graines

increased over the past decade, reaching 350,000 tons today, mainly to the USA and Europe. Traditional uses for medicine and foods account for 50%, refining for specialty fats 45%, and cosmetics (refined or not) 5% of production (Lovett personal communication).

In order to improve the sustainable management of this species, to launch prebreeding programs and precisely assess the potential of different geographical origins for marketing, it is now becoming crucial to improve our knowledge of the diversity of the chemical content of nuts.

Initial evaluations of Shea nut chemical composition throughout the natural range of *Vitellaria paradoxa* have shown great diversity particularly for fat content and fatty acid composition (4, 5). Those studies revealed fat content values ranging from 22.3% to 52.8%. Relative fatty acid compositions were found to vary, particularly for oleic acid (C18:1), ranging from 37.1% to 62.1%, and stearic acid (C18:0), from 25.6% to 50.2%. It was suggested that the variation observed in these initial investigations was due to the two subspecies, but also to biodynamic variations related to temperature and annual rainfall.

However, these first evaluations of Shea nut chemical diversity were based on an unbalanced and restricted collection of material. Consequently, the results may not reflect a statistically accurate assessment of actual *Vitellaria paradoxa* variability within its natural range. Indeed, some geographical origins studied were represented by 29 samples, while others by only one.

Fresh assessment of chemical content variability, based on significant and representative sampling, is therefore needed. Classical chemical tools such as solvent extraction and gas chromatography (GC) (6, 7) have shown their efficiency in determining the chemical composition of plant material, but they are time-consuming and expensive. In this context, there is now a definite need to develop a rapid and accurate method.

Near-infrared spectroscopy (NIRS) is a very efficient method for high-throughput screening of plant materials for their chemical characteristics. This indirect method is based on vibrational properties of organic molecule chemical bonds and their interactions with infrared radiation. The NIR absorption spectrum is therefore correlated with a sample's chemical composition (8). The NIRS method can therefore be used to predict the value of a chemical trait in a sample, but requires the development of an accurate and robust calibration equation. To that end, this method requires prior characterization of a large number of samples uniformly covering the range of variability of the studied trait (9). Compared to conventional time-consuming biochemical methods, NIRS has already proven its usefulness for estimating various parameters in diverse natural and agricultural products (10). Indeed, several studies have demonstrated NIRS efficiency in characterizing fat content and fat profiles in fruit from various species such as peanut, hazelnut and sesame (11–14).

In this paper, we demonstrate the efficiency of near-infrared spectroscopy in characterizing fat profiles for nuts from Shea trees of different geographical origin and we propose calibration equations. This technique provides producing countries with an opportunity for the high-throughput characterization of large numbers of Shea nut samples. Moreover, use of the NIRS technique will be helpful in phenotype/genotype association studies and breeding programs.

MATERIALS AND METHODS

Materials. The sampling strategy was designed to ensure maximum coverage of the range of variation in fat composition. As suggested by previous studies (4, 5), variation was assumed to be linked to the difference between subspecies *paradoxa* and *nilotica*, and to environmental gradients. Consequently, within each subspecies area, samples were collected according

Table 1. Description of the 17 Harvested Sites

site index	site	GPS	altitude (m)	N ^a	
				2007	2008
Senegal					
1	Kenioto	12.57N; 12.16W	161		12
2	Samecoutha	12.60N; 12.13W	126		18
3	Saraya	12.83N; 11.75W	180		10
Mali					
4	Nafégué	10.51N; 5.98W	344	35	40
5	Mperesso	12.28N; 5.33W	340	34	37
6	Daelan	13.25N; 4.99W	282	34	37
7	Tori	13.61N; 3.72W	377		35
8	Sassambourou	14.31N; 3.51W	392		35
Burkina					
9	Titao	13.72N; 2.16W	336		18
10	Guibare	13.07N; 1.61W	303		21
Ghana					
11	Kawampe	8.43N; 1.56W	125		35
12	Tolon	9.43N; 1.00W	154		35
13	Kulbia	10.83N; 0.96W	206		34
Uganda					
14	Katakwi	1.82N; 33.99E	1100	34	29
15	Pader	2.80N; 33.31E	1031	31	25
16	Moyo	3.62N; 31.64E	863		16
17	Uleppi-Arua	3.02N; 30.90E	1200		19

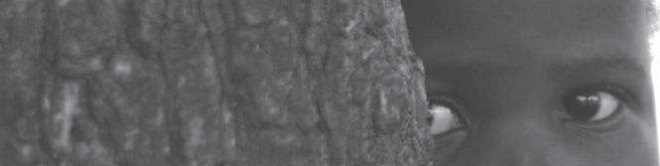
^a N: number of sampled trees. Total: 624.

to north–south dunes in order to increase environmental variations (rainfall and temperature decreases from North to South).

Samples were collected under uniform conditions over two years (2007 and 2008) in four West African countries, Senegal, Mali, Burkina Faso and Ghana for subspecies *paradoxa*, and one East African country, Uganda, for subspecies *nilotica*. Within each country, different sites were sampled based on a rainfall and temperature gradient. A total of 624 trees (GPS located) were sampled at 17 sites: Senegal (Keniato, Samecouita and Saraya), Mali (Nafégue, Mperesso, Daelan, Tori and Sassambourou), Burkina Faso (Titao and Guibare), Ghana (Kawampe, Tolon and Kulbia) and Uganda (Katakwi, Pader, Moyo and Uleppi-Arua). A sampling description is given in Table 1, and the distribution of sites is illustrated in Figure 1. On average, 30 mature fruits (ready to fall to the ground) were collected per tree and pooled. Postharvest treatment, including depulping and drying (3 days at 60 °C), was carried out on site. The dried nuts were sent to the CIRAD laboratory in Montpellier (France). On receipt, nuts were oven-dried for two days at 60 °C in order to stabilize moisture content, and then they were stored at room temperature prior to analysis.

Sample Preparation. Unshelled Shea nuts were first ground in a "Vorwerk Thermomix Robot". Raw powders were frozen at -20 °C and reground in a "SEB Valentin blender" in order to obtain a final particle size between 0.5 and 0.8 mm. The final powder samples were stored at -20 °C.

Laboratory Analyses. For each sample, moisture content (MC) was assessed by gravimetric analysis after drying at 103 °C in an oven (Chopin) for 16 h. Fat content (FC) was solvent-extracted (petroleum ether) from powders using a semiautomatic Soxtec 2050 extractor (FOSS Analytical, A/S Slangerupgade 69, DK-3400 Hillerød, Denmark) according to the manufacturers' instructions. In a preliminary study (data not shown), we compared our extraction procedure with conventional fat extraction method (Soxhlet with hexane as the solvent over 16 h) and observed no significant differences. After gravimetric quantification, extracted oils were stored at -20 °C for further chemical analyses. Fatty acid (FA) profiles were obtained according to the protocol described by Tchobo (6): after esterification of the oil using sodium methylate, FA profiles were determined by gas chromatography using a Thermo Focus (Thermo Fisher Scientific, 81 Wyman Street, Waltham, MA 02454) GC with a



Chapitre III : Variabilité des constituants chimiques des graines

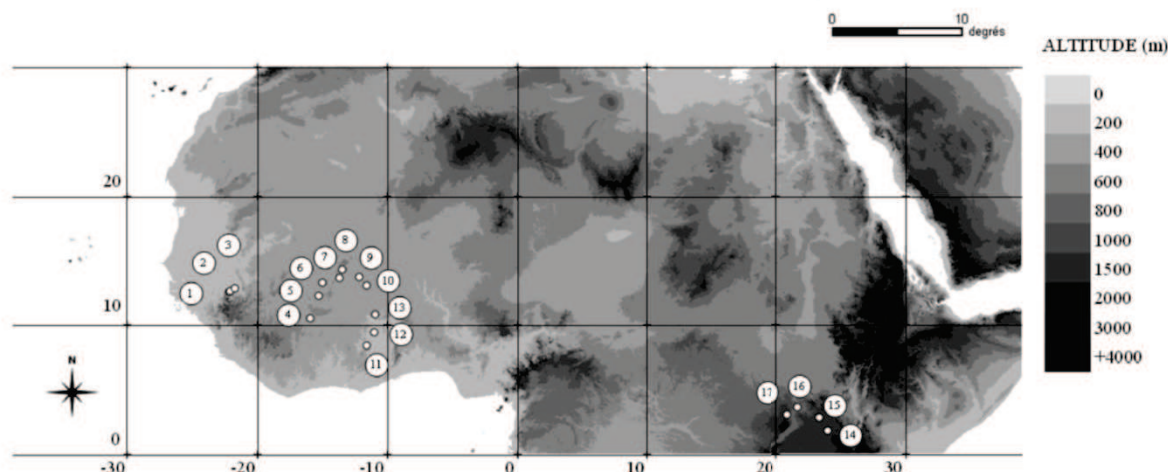


Figure 1. Location of the 17 harvested sites.

CP SIL 88 (highly substituted cyanopropyl phase) column (Varian, Inc., 3120 Hansen Way, Palo Alto, CA 94304-1030). The chromatographic conditions were as follows: the injector temperature was 250 °C and that of the FID detector 270 °C. The oven temperature settings were as follows: 150 °C to 225 at 5 °C/min, and then held at 225 °C for 2 min. Vector (He) gas flow was 1 mL/min with a splitting ratio of 1:80. The injection volume was 1 µL. For all laboratory measurements the standard error of the laboratory (SEL) was estimated as the standard deviation from 10 replicates of a standard Shea sample.

For each parameter, comparisons between sites and subspecies were carried out using one-way analyses of variance (ANOVA) with Student-Newman-Keuls SNK multiple pair comparison tests at the 5% level (XLstat software, Addinsoft, 40 rue Darnémont, 75018 Paris, France). Outlier samples were detected using the Grubbs test (15) procedure at the 5% level (Statgraphics Centurion XV, Statpoint Technologies, Inc., 560 Broadview Avenue, Suite 201, Warrenton, VA 20186).

NIR Spectrum Acquisition. A NIRS 6500 monochromator (Foss NIRSystems, Silver Spring, MD) was used to scan reflectance from 400 to 2500 at 2 nm intervals, using ring cups (50 mm in diameter) with about 3 g of fine Shea nut powder. Data were saved as the average of 32 scans and stored as $\log(1/R)$, where R was the reflectance at each wavelength and 1 the reflectance of a standard ceramic reference. Spectra were acquired randomly, each sample being measured twice, and the average spectrum was stored. Statistical analyses were performed using WinISI II software (Infrasoft International, Port Matilda, PA), JUMP 7.01 (SAS Institute Inc., Cary, NC), Statgraphics Centurion XV (StatPoint Inc., Warrenton, VA) and XLstat software (Addinsoft, Paris, France).

Spectrum Pretreatment. Spectra were mathematically corrected for light scattering by using the standard normal variate and detrend correction (16, 17). The second derivative was calculated on five data points and smoothed using Savitzky and Golay polynomial smoothing on five data points (18).

Principal Component Analysis. Prior to calibration development, a principal component analysis (PCA) was used to extract relevant information from the spectral matrix ($n = 624$). The generalized Mahalanobis distance (H) was calculated on the extracted PCs for each sample. This statistical distance is useful for defining boundaries of the population and a similarity index between spectra (19).

This enabled us to discard 22 outlier samples with a Mahalanobis distance $H > 3$. These samples had been tagged as moldy on arrival at the laboratory.

NIR Calibration Development. In order to assess the performance of the predictive equations, the remaining 602 PCA samples were split into a calibration subset (cal) and a validation subset (val). The validation set was created by randomly selecting 10% of the 602 samples including 2 samples from Burkina, 8 from Ghana, 27 from Mali, 1 from Senegal and 22 from Uganda. The calibration set comprised the remaining 542 samples.

Calibration equations for the parameters were constructed with the calibration subset (542 samples) using the modified partial least-squares regression (mPLS) (20) algorithm of WinISI software (9).

Calibration statistics included the following parameters: standard deviation (SD), coefficient of determination (R^2), standard error of calibration (SEC), and standard error of cross-validation (SECV). Cross-validation was used during calibration development in order to select the optimum number of latent variables and to minimize overfitting of the equations (21, 22). For SECV estimation, 25% of the samples (randomly selected) were predicted using a calibration model developed with the other 75%. SECV estimation was repeated four times and the average calculated. In addition to R^2 , the ratio of performance to deviation (RPD = $SD/SECV$) was used to evaluate the general quality of the fit obtained for each equation. Unlike SEC and SECV, RPD is independent of parameter units and can therefore be compared between parameters (23). The Student (t) test was used to identify t -outlier samples during calibration development. Outlier detection was based on the standardized residuals (= $error/SECV$) with a cutoff of 2.5. Two passes of outlier elimination were used (22).

The standard error of prediction (SEP), corresponding to the standard deviation of residuals, was estimated by predicting the validation subset using a model developed on the calibration subset. The ratio performance to deviation of prediction (RPD_p) was also calculated as $RPD_p = SD_{val}/SEP$ (where SD_{val} was the standard deviation of validation samples). The quality of the fits between wet chemistry values (moisture, fat and FA relative percentages) and NIRS-predicted values was evaluated from the linear regression slope, the R^2 and the bias.

RESULTS AND DISCUSSION

Chemical Analysis. The 602 samples were analyzed in the laboratory for their moisture and fat content, and 599 extracted butter fatty acid profiles were determined by GC (Table 2).

Moisture content ranged from 2.25% to 8.37%, with an average value of 4.48%. Data dispersion was rather small ($SD = 0.89\%$). Low moisture content values avoided lipase activity and free fatty acid formation (24). Five samples presented extreme values (Grubbs test; level 5%) with a moisture content ranging between 7.49% and 8.37%. No particular defect was noticed on receipt of these samples. Their corresponding NIR spectra were not atypical (H values lower than 3), and the calibration process did not highlight these samples as outliers. This relatively high moisture content had no incidence on calibration development and was probably the consequence of incomplete drying.

In order to avoid bias in spectral pattern interpretation, an analysis of variance (ANOVA) was performed to ensure that sample origins had no effect on moisture content values. The ANOVA concluded that there was no significant difference (level $R = 5\%$) between countries for average moisture content values. Fat content was equal to 49.66% on average (dry matter basis) with a SD of 5.03% corresponding to relatively low dispersion

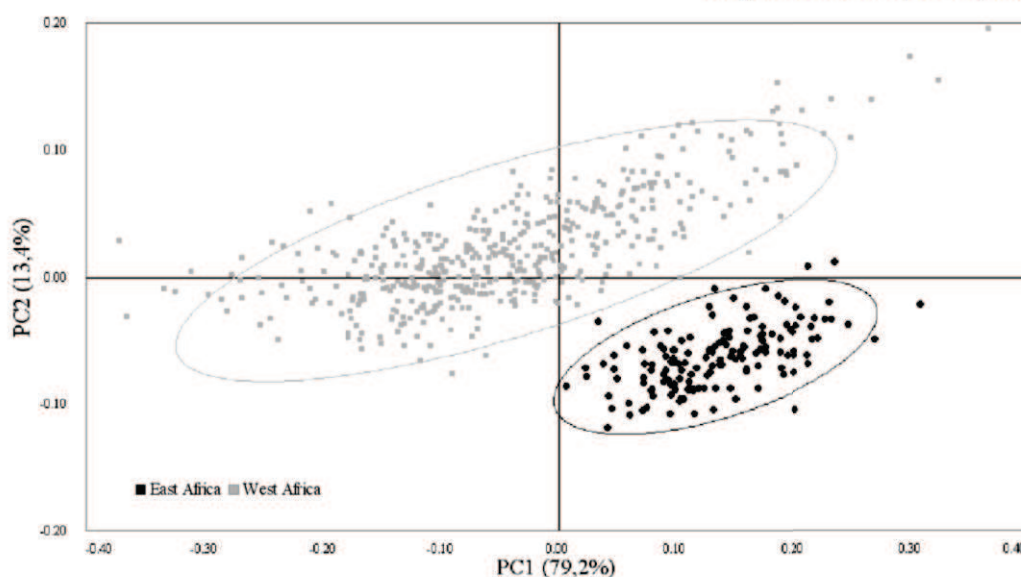


Figure 4. Scatter plot of the 602 sample scores for the first two principal components (95% confidence ellipses).

low (0.41% and 0.31%) and detected in 565 and 509 samples respectively. Linoleic acid was detected in all butters, ranging from 3.22% to 13.0%, with an average of 7.13%. The relative content of linoleic acid was fairly similar for the five countries, with a minimum value (6.59%) for Senegalese butters and a maximum (7.25%) for Malian butters. Arachidic acid ranged from 0.46% to 1.83%, with an average of 1.20%. Arachidic acid was found to be lower in East Africa (0.98% in Uganda) and higher in West African countries (1.16% in Senegal, 1.17% in Ghana, 1.26% in Burkina and 1.33% in Mali).

The average ratio of saturated fatty acids (SFA) to unsaturated (USFA) was 0.79. Ugandan Shea butter showed the lowest ratio (0.55) while for West African butters the ratio ranged from 0.85 (Mali) to 0.92 (Burkina).

cis-Vaccenic, linoleic and arachidic acids followed normal distributions (figures not shown). Stearic and oleic acid distributions were bimodal due to West and East African samples (Figure 3). Linolenic acid distribution was also bimodal with Malian butters apart from the others.

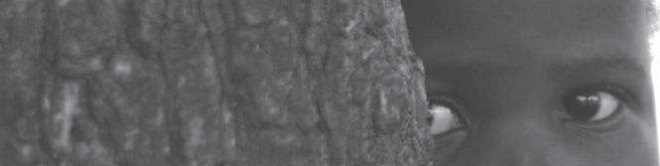
According to the ANOVA-SNK test (level 5%), significant differences were found for stearic and oleic acid content between Uganda (subsp. *nilotica*) and West Africa (subsp. *paradoxa*). Within West Africa, for those FAs, Mali and Ghana were found to be significantly different. Palmitic acid content was similar for Malian and Burkinese butters, and different for other countries. *cis*-Vaccenic acid content was significantly different between Uganda and Mali. Linoleic acid content was similar for Senegal and Burkina, and different for the other countries. Linolenic acid content was similar for Ugandan and Senegalese butters, and for Ghanaian and Burkinese butters, while Malian butters were different from the others. Finally, only butters from Ghana and Senegal displayed a similar relative arachidic acid content.

Spectral Data Matrix: Principal Component Analysis. The first three principal components (PCs) extracted from a PCA carried out on the 602 samples explained 79.2%, 13.2%, and 3.5% of total inertia respectively. Only three samples presented H distances, calculated on 32 PCs, over 3: one from Mali ($H = 3.1$), one from Uganda ($H = 3.2$) and one from Senegal ($H = 3.5$). In terms of fat and moisture content these samples were in the normal range, and no special information was registered for them, therefore based on their relatively low H values they were kept in the database for the rest of the study.

A scatter plot of the first two PCs (Figure 4) showed a sharp discrimination between West and East African Shea nuts. Based on previous results (4), the observed discrimination was suspected of being due to differences in butter FA profiles between West and East African Shea nuts. This hypothesis was confirmed through the significant correlations (Pearson test, 5% level) calculated between PCs and fatty acids (e.g., correlation between PC1 and stearic acid $r = -0.83$; PC1 and oleic $r = 0.79$). Additionally, we found a significant correlation between PC2 and fat content ($r = -0.86$).

The scatter plot of eigenvector coefficients (loadings) versus wavelength confirmed the origin of the discrimination (Figure 5). Indeed, the highest coefficients were associated with the wavelength corresponding to absorption bands of chemical bonds specific to fat, such as the C-H from $-CH_2$ group second overtone stretching band (1214 nm), the $-CH_2$ stretch combination band (1396 nm), the C-H ($-CH_2$ group) first overtone stretching band (1724 nm) and the characteristic C-H ($-CH_2$ group) stretching and deformation combination band (2308 nm) (10, 26). The profiles for loadings 1 and 2 were similar for major absorption bands, the absorption band due to H-OH second overtone (1916 nm) being taken into account by PC2. Assignment of the major Shea nut NIR absorption bands is summarized in Table 3. Moreover, the loading profiles were close to the pure butter spectrum, which accounts for a large share of the Shea nut spectral fingerprint. This was illustrated by comparing the Ugandan Shea nut spectrum and the corresponding extracted butter spectrum (Figure 6).

NIRS Calibration. Of the 602 spectra, the 542 samples used for calibration and the 60 randomly selected samples for validation were representative of the variation, in terms of SD and range (Table 4) for each constituent. Fat and stearic acid distributions for validation and calibration sets are shown in Figure 7. These histograms illustrate the distribution of validation samples with respect to the variation of the population. One validation sample had an extreme value (26.96%) for fat content (lower than the lowest calibration sample whose fat content was 28.44%); keeping this sample in the validation set allowed us to test our fat model in extrapolation. Modified PLS regressions (mPLS) were performed using the calibration set and reference laboratory data. Constituents were then predicted for the validation set using the mPLS models.



Chapitre III : Variabilité des constituants chimiques des graines

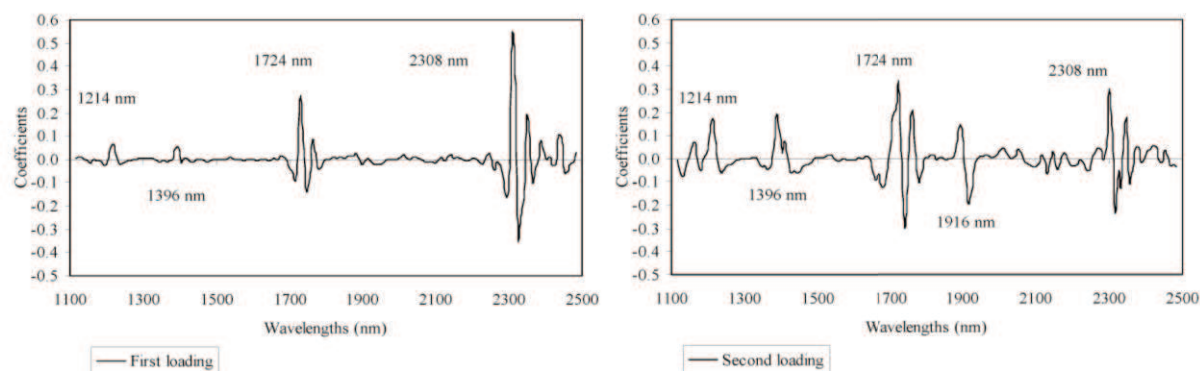


Figure 5. The first two loadings as a function of wavelength.

Table 3. Assignment of Major Shea Nut NIR Absorption Bands

wavelength (nm)	bond type	vibration mode	structure
1160	C-H	stretching 2nd overtone	CH ₃
1172	C-H	stretching 2nd overtone	HCO CH
1214	C-H	stretching 2nd overtone	-CH ₂
1396	-CH ₂	stretching	-CH ₂
1450	O-H	stretching 1st overtone	H ₂ O
1724	C-H	stretching 1st overtone	-CH ₂
1916	O-H	stretching and deformation combination	H ₂ O
2100- 2200	Cd C	stretching and deformation combination	HCO CH
	dis d C- H and Cd C	stretching combination	HCO CH
2308	C-H	stretching and deformation combination	-CH ₂

Model prediction accuracy was evaluated through RPD_p: values below 1.4 corresponded to unusable calibration, RPD_p values between 1.4 and 1.7 corresponded to a calibration usable for rough screening, and values higher than 4 indicated that the calibration could be used for quality control analyses (10, 27). Calibration statistics and validation statistics are reported in Table 4.

Moisture Content. Calibration for MC gave both R² and R_p² of 0.95. For this model, RPD_p was 4.45 and SEP was 0.23% with a regression slope of 0.96 between laboratory and NIR-predicted values (Figure 8a). Compared to calibrations obtained in the literature for MC (28), this model appeared less efficient. However, this was due to sample preparation (nuts were dried), which led to dissymmetric distribution of MC values: 83% of samples contained less than 4% moisture. In any event, based on model performances, MC NIR calibration can be applied to Shea nuts for routine analysis. To prevent lipid degradation due to hydrolysis of triacylglycerol, commercial nuts must have a moisture content below 9% (29), thus our model was efficient enough for its control.

Fat Content. The model developed for fat quantification was efficient: R² and R_p² were equal to 0.96 and 0.97, and RPD_p to 5.61. The regression slope between laboratory and NIR-predicted values was 1.02 (Figure 8b). The estimated SEP was 1.05%, enabling Shea nut fat content determination with an accuracy of (2.06%. The model obtained was in accordance with results in the literature related to high fat content products (30). Our calibration was usable for quality control of Shea nut fat content in producing countries for commercial purposes.

cis-Vaccenic Acid. The calibration developed for cis-vaccenic acid was poor with R² = 0.48. This result was confirmed by the estimation of the R_p² on the validation set (R_p² = 0.23). The RPD_p

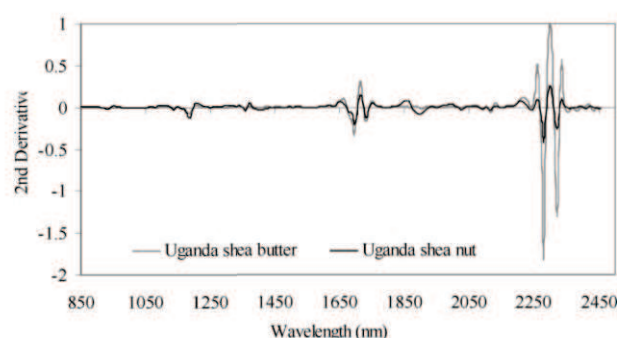


Figure 6. Spectra of Shea nut powder (Uganda) and corresponding extracted butter.

observed for the cis-vaccenic acid model was about 1.2, corresponding to an unusable calibration. This was probably due to the low cis-vaccenic acid contents observed (< 0.9%) which were close to the NIRS detection limit.

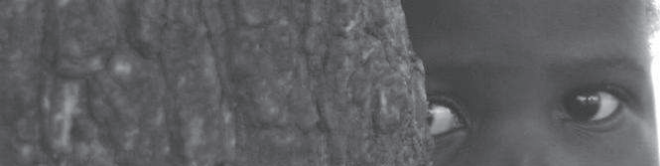
Palmitic Acid. The palmitic acid calibration had a low R_p² (0.53) for the validation set compared to R² (0.61) of the calibration set. This was due to an overestimation of the R² coefficient resulting from extreme values in the calibration set. RPD_p obtained with the model (1.3) also corresponded to an unusable calibration. The lack of correlation between the NIR fingerprint and chemical data resulted from the low variability of palmitic acid content which ranged from 3.31% to 5.62% for 83.8% of samples.

Linoleic Acid. For linoleic acid, the R² value for calibration (0.63) was similar to R_p² (0.68). This reflected a relatively efficient model for both calibration and prediction. With an RPD_p equal to 1.71, the calibration developed for linoleic acid could therefore be used for rough screening.

Arachidic Acid. For arachidic acid, the model developed was similar to the linoleic acid model with R² = 0.63, R_p² = 0.57 and RPD_p = 1.50. To conclude, as for linoleic acid, this model could only be applied for very rough screening.

Linolenic Acid. The calibration model for linolenic acid seemed fairly good, as for arachidic and linoleic acids. However, validation revealed the poor performance of the model, with a very low R_p² (0.39) and RPD_p (1.30). This result highlighted the importance of using a validation set covering the whole range of values. Indeed, the R² observed for linolenic acid was artificially high, due to a bimodal distribution of values for the calibration samples.

Stearic and Oleic Acids. The R² values for calibrations were close to 1 for stearic acid (0.96) and oleic acid (0.98). High R² coefficients correspond to a very good data fit. These results were confirmed by estimating R_p² on the validation set (stearic acid



Chapitre III : Variabilité des constituants chimiques des graines

Table 4. Descriptive Statistics for Calibration Subsets, Validation Subsets and NIR Equations^a

	MC	fat	palmitic	stearic	oleic	cis-vaccenic	linoleic	linolenic	arachidic	ratio ^b
Calibration N = 542										
mean	4.46	49.68	4.24	38.25	48.44	0.40	7.15	0.30	1.20	0.79
range	6.12	31.22	5.73	25.32	23.58	0.82	8.47	0.63	1.37	0.76
SEL	0.1	0.79	0.19	0.98	0.66	0.09	0.34	0.12	0.05	0.02
SD _{cal}	0.84	4.77	0.66	5.68	5.21	0.15	0.90	0.12	0.19	0.17
SEC	0.18	1.00	0.41	1.06	0.74	0.11	0.55	0.07	0.12	0.03
R ²	0.95	0.96	0.61	0.96	0.98	0.48	0.63	0.66	0.63	0.97
SECV	0.20	1.08	0.46	1.14	0.81	0.11	0.62	0.08	0.13	0.04
RPD	4.22	4.43	1.43	4.97	6.47	1.32	1.46	1.54	1.53	4.25
Validation N = 60										
mean	4.44	50.40	4.49	36.69	49.68	0.41	7.28	0.31	1.17	0.76
range	4.65	31.38	3.06	24.16	23.15	0.69	9.11	0.49	0.87	0.79
SD _{val}	1.02	5.88	0.69	7.43	7.11	0.19	1.33	0.13	0.21	0.23
SEP	0.23	1.05	0.53	1.19	0.90	0.16	0.78	0.10	0.14	0.04
R _p ²	0.95	0.97	0.45	0.98	0.99	0.23	0.68	0.39	0.57	0.97
RPD _p	4.45	5.61	1.30	6.26	7.91	1.19	1.71	1.30	1.50	5.72

^a MC: moisture content. N: number of samples. SEL: standard error of laboratory. SD: standard deviation for calibration subset. SEC: standard error of calibration. R²: coefficient of multiple determination. SECV: standard error of cross validation. RPD (ratio performance to deviation) = SD_{cal}/SEC. SD_{val}: standard deviation for validation subset. SEP: standard error of prediction; R_p²: coefficient of multiple determination for prediction. RPD_p (ratio performance to deviation for prediction) = SD_{val}/SEP. ^b Ratio saturated to unsaturated fatty acids.

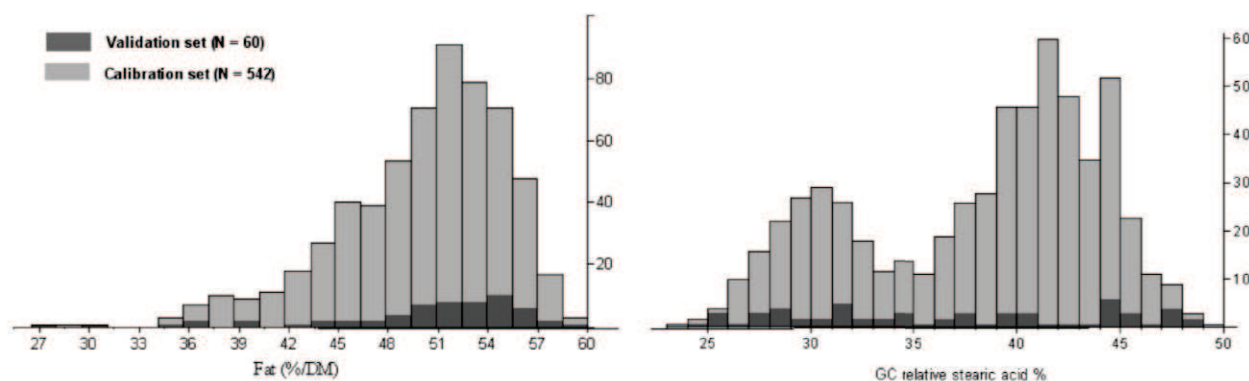


Figure 7. Distributions of fat content and stearic acid % for calibration and validation subsets.

0.98 and oleic acid 0.98). The RPD_p values obtained for stearic and oleic models were 6.26 and 7.91 respectively. Stearic and oleic acids displayed bimodal distributions due to eastern and western butter specificities, however the fit was efficient for each sub-population, resulting in an overall accurate calibration. SEP values observed on the validation set for stearic and oleic acids were 1.19% and 0.90% respectively. The regression slopes between GC values and NIR-predicted values were 1.04 and 1.06 respectively (Figures 8c and 8d). Both models were set up with a similar number of PLSTerms (10 and 11), though the oleic acid calibration performed better. Oleic acid contains a double C-bond (Cd C) with a specific signature between 2100 and 2200 nm (30). The slight difference between the stearic and oleic models was therefore probably due to this specificity resulting in a higher correlation between oleic acid GC values and NIR data. Furthermore, for regression coefficients, higher correlations were found at 1160 nm and 1172 nm, corresponding to double bond-related wavelengths (Table 3) and in the 2140- 2180 nm region assigned to the combination of cis d C- H and Cd C stretching, - CH₂ asymmetric stretching and Cd stretching of the - HCd CH-structure (31).

Saturated vs Unsaturated Acids. Calibrations developed for total saturated fatty acids and total unsaturated fatty acids (mono- and polyunsaturated) led to similarly efficient models,

with a RPD_p of 5.72 (Table 4). This performance was not surprising due to the high quality of individual calibrations obtained for oleic and stearic acids which represented 86% of unsaturated FA and 87% of saturated FA acids in Shea butter.

Industrial and Research Prospects. Obtaining a precise description of Shea nut quality for sustainable development, research and industry is becoming ever more critical. In this study, we proposed NIR models for accurate prediction of Shea nut stearic and oleic acid composition, fat and water content. Transferring these novel tools to producing countries will provide them with the opportunity to control and promote the quality of their production. In addition, for the food and cosmetics industries, this application will enable early selection of products according to their end use. Accurately predicting the saturated:unsaturated FA ratio is relevant for Shea butter use as a cocoa butter equivalent (CBE) in the chocolate industry.

In terms of research, understanding the environmental and genetic basis of variation patterns in the fat composition of Shea nuts is an important issue. Studies being developed (data not published) reveal a specific genetic variation pattern for a gene strongly involved in the variation of the relative percentages of oleic and stearic acids, and by extension the saturated:unsaturated fatty acid ratio (32, 33). To that end, our NIR models

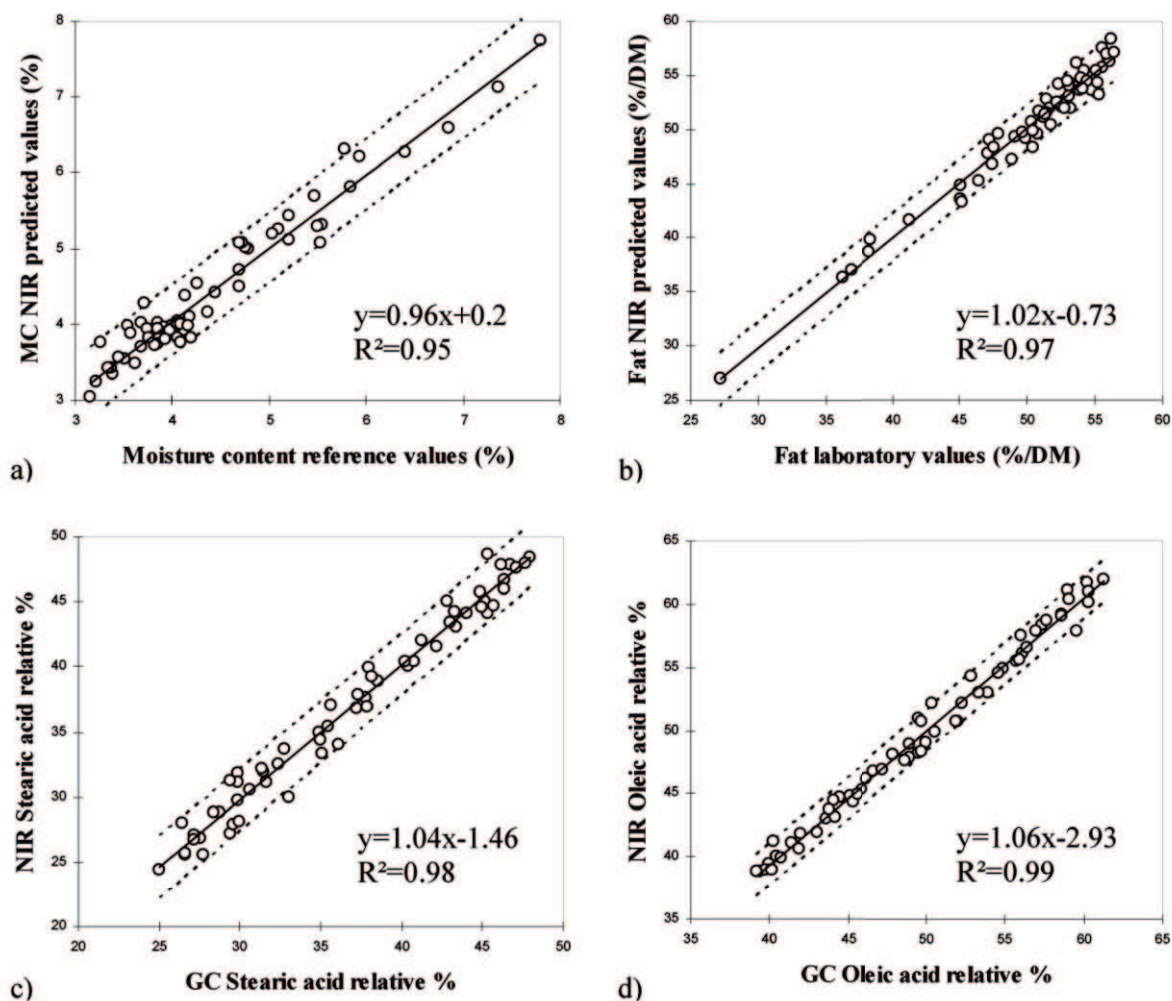


Figure 8. Scatter plots of laboratory values versus NIR-predicted values for the validation subset (95% confidence interval). (a) Moisture content as a %. (b) Fat content as a % of dry matter (DM). (c) Relative stearic acid composition as a %. (d) Relative oleic acid composition as a %.

enable us to achieve an accurate, quick and cheap determination of a large number of samples needed for quantitative genetics studies.

In this paper, based on a large collection of samples, we confirmed a differentiation in Shea nut fat composition correlated to geographical origin. The performance of the calibrations obtained indicated that Shea nut moisture and fat contents could be determined with acceptable accuracy using NIR spectroscopy methods. The good calibration performance for relative stearic and oleic acid percentages indicated their suitability for fine determination. NIR calibration enables fine characterization of fat profiles in terms of saturated and unsaturated fatty acids. We demonstrated that NIR allows rapid (less than one minute per sample) nondestructive and reliable determination in one shot of moisture, fat, stearic and oleic acid contents of Shea nuts. Thus our models can be applied for high-throughput characterization of Shea nut quality. Transferring this technology will enable a Shea nut quality control and traceability survey. Moreover, our NIR models make it possible to carry out further quantitative genetics investigations.

ACKNOWLEDGMENT

The laboratory work was conducted at the CIRAD laboratory in Montpellier France, and chemical analyses were carried out by F. Allal, L. Millet and N. Forestier-Chiron. Field activities were conducted by scientists from African countries and from CIRAD,

and we thank all our co-workers for their help in fruit collection. We thank Peter Biggins and Guy Self, from CIRAD, for revising the English.

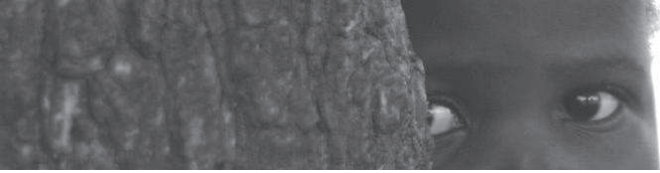
LITERATURE CITED

- (1) Hall, J. B.; Aebischer, D. P.; Tomlinson, H. F.; Osei-Amaning, E.; Hindle, J. R. *Vitellaria paradoxa: A Monograph*; School of Agricultural and Forest Sciences, University of Wales: Bangor, U.K.: 1996; p 105.
- (2) Neumann, K.; Kahlheber, S.; Uebel, D. Remains of woody plants from Saouga, a medieval West African village. *Veget. Hist. Archaeobot.* 1998, 7, 57–77.
- (3) Lovett, P.; Haq, N. Evidence for anthropic selection of the Shea nut tree (*Vitellaria paradoxa*). *Agroforestry systems* 2000, 48, 273–288.
- (4) Maranz, S.; Wiesman, Z.; Bisgaard, J.; Bianchi, G. Germplasm resources of *Vitellaria paradoxa* based on variations in fat composition across the species distribution range. *Agrofor. Syst.* 2004, 60, 71–76.
- (5) Di Vincenzo, D.; Maranz, S.; Serraiocco, A.; Vito, R.; Wiesman, Z.; Bianchi, G. Regional Variation in Shea Butter Lipid and Triterpene Composition in Four African Countries. *J. Agric. Food Chem.* 2005, 53, 7473–7479.
- (6) Tchobo, F.; Natta, A.; Barea, B.; Barouh, N.; Piombo, G.; Pina, M.; Villeneuve, P.; Soumanou, M.; Sohounhloue, D. Characterization of *Pentadesma butyacea* sabin Butters of Different Production Regions in Benin. *J. Am. Oil Chem. Soc.* 2007, 84, 755–760.
- (7) Buchgraber, M.; Ulberth, F.; Anklam, E. Cluster Analysis for the Systematic Grouping of Genuine Cocoa Butter and Cocoa Butter Equivalent Samples Based on Triglyceride Patterns. *J. Agric. Food Chem.* 2004, 52, 3855–3860.

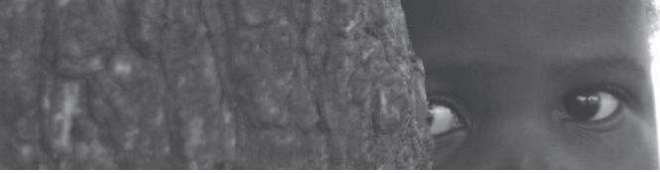
Article

- (8) Pasquini, C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* 2003, 14, 198–219.
- (9) Shenk, J. S.; Westerhaus, M. O. Population Definition, Sample Selection, and Calibration Procedures for Near Infrared Reflectance Spectroscopy. *Crop Sci.* 1991, 31, 469–474.
- (10) Williams, P.; Norris, K. Near-Infrared Technology: in the Agricultural and food industries, 2nd ed.; American Association of Cereal Chemists, Inc: USA, 2001; p 330.
- (11) Fox, G.; Cruickshank, A. Near infrared reflectance as a rapid and inexpensive surrogate measure for fatty acid composition and oil content of peanuts (*Arachis hypogaea* L.). *J. Near Infrared Spectrosc.* 2005, 13, 287–291.
- (12) Sato, T. New Estimation Method for Fatty Acid Composition in Oil Using Near Infrared Spectroscopy. *Biosci., Biotechnol., Biochem.* 2002, 66, 2543–2548.
- (13) Sato, T.; Kawano, S.; Iwamoto, M. Near infrared spectral patterns of fatty acid analysis from fats and oils. *J. Am. Oil Chem. Soc.* 1991, 68, 827–833.
- (14) Sato, T.; Maw, A.; Katsuta, M. NIR reflectance spectroscopic analysis of the FA composition in sesame (*Sesamum indicum* L.) seeds. *J. Am. Oil Chem. Soc.* 2003, 80, 1157–1161.
- (15) Grubbs, F. E. Procedures for Detecting Outlying Observations in Samples. *Technometrics* 1969, 11, 1–21.
- (16) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* 1989, 43, 772–777.
- (17) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Letter: Correction to the description of Standard Normal Variate (SNV) and De-Trend (DT) transformations in Practical Spectroscopy with Applications in Food and Average Analysis-2nd Edition. *J. Near Infrared Spectrosc.* 1993, 1, 185–186.
- (18) Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 1964, 36, 1627–1639.
- (19) Shenk, J. S.; Westerhaus, M. O. Routine Operation, Calibration Development Manual; Infrasoft International: Silver Spring, MD, 1996.
- (20) Wold, H. Non-linear iterative partial least squares (NIPALS) modelling. Some current developments. In *Multivariate Analysis*; Krishnaiyah, P. R., Ed.; Academic Press: New York, 1973; Vol. III, pp 387–407.
- (21) Wold, S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 1978, 20, 397–405.
- (22) Burns, D. A.; Ciurczak, E. W. E. Handbook of Near-Infrared Analysis; Marcel Dekker, Inc.: New York, 1992; p 680.
- (23) Sinnaeve, G.; D., P.; Agneessens, R.; Biston, R. The use of near infrared spectroscopy for the analysis of fresh grass silage. *J. Near Infrared Spectrosc.* 1994, 2, 79–84.
- (24) Verger, R.; Riviere, C. Les enzymes lipolytiques: une étude cinétique. *Rev. Fr. Corps Gras.* 1987, 1, 7–13.
- (25) Lovett, P. N.; Haq, N. Diversity of the Sheanut tree (*Vitellaria paradoxa* C.F. Gaertn.) in Ghana. *Genet. Resour. Crop Evol.* 2000, 47, 293–304.
- (26) Osborne, B.; Fearn, T.; Hindle, P. Practical NIR spectroscopy: with applications in food and beverage analysis; Longman Scientific & Technical: Harlow, England, 1993; p 227.
- (27) Kovalenko, I.; Rippke, G.; Hurburgh, C. Measurement of soybean fatty acids by near-infrared spectroscopy: Linear and nonlinear calibration methods. *J. Am. Oil Chem. Soc.* 2006, 83, 421–427.
- (28) Guyot, B.; Davrieux, F.; Manez, J. C.; Vincent, J. C. Determination of caffeine and dry matter content by Near Infrared spectrometry. Applications in raw Robusta coffee and roasted coffees. *Café Cacao Thé* 1993, 37, 53–64.
- (29) ProKarité Regional Technical Committee on Africa regional standard for shea kernel: www.ProKarité.org.
- (30) Panford, J.; deMan, J. Determination of oil content of seeds by NIR: Influence of fatty acid composition on wavelength selection. *J. Am. Oil Chem. Soc.* 1990, 67, 473–482.
- (31) Sohn, M.; Kim, Y.; Vines, L. L.; Kays, S. W. Near infrared analysis of lipid classes in processed cereal products. *J. Near Infrared Spectrosc.* 2009, 17, 127–133.
- (32) Shanklin, J.; Somerville, C. Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. *Proc. Natl. Acad. Sci. U.S.A.* 1991, 88, 2510–4.
- (33) Somerville, C.; Browse, J.; Jaworski, J. G.; Ohlrogge, J. B. Lipids. In *Biochemistry and Molecular Biology of Plants*; Buchanan, B., Jones, R., Gruissem, W., Eds.; American Society of Plant Physiologists: Rockville, MD, 2000; Chapter 10.

Received for review February 1, 2010. Revised manuscript received May 20, 2010. Accepted May 20, 2010. This study was one of the tasks of the INCO INNOVKAR project funded by the European Union for 'Innovative Tools and Techniques for Sustainable Use of the Shea Tree in the Sudano-Sahelian zone'.



Chapitre IV : Étude du polymorphisme des gènes candidats



Chapitre IV : Étude de la variabilité de gènes candidats



1 INTRODUCTION

1.1 Comment détecte-t-on la signature de la sélection naturelle dans les génomes ?

Au cours de l'introduction générale de cette thèse, nous avons posé les bases d'une réflexion sur l'évolution adaptative et l'impact de la sélection naturelle sur les populations. Cependant, si des preuves d'évolution adaptative ont été rapportées dans la littérature depuis les travaux de Kettlewell sur la phalène du bouleau (Kettlewell, 1958), les bases génétiques de l'adaptation des populations naturelles demeurent encore mal comprises (Orr, 2005).

Selon la théorie neutre de la sélection (Kimura, 1968, 1985), la plupart de la variation génétique au sein des espèces est sélectivement neutre, c'est à dire qu'elle n'affecte pas la « fitness » des organismes. Sous ce principe, si l'histoire démographique d'une population affecte tout le génome de la même façon, modifiant aléatoirement la fréquence d'allèles ne fournissant pas un avantage adaptatif pour les organismes qui les portent, la sélection s'effectue de manière locus-spécifique (Cavalli-Sforza, 1966). Dans ce contexte, la recherche de traces de sélection au sein des génomes et des gènes permet d'émettre des hypothèses cruciales pour la compréhension de l'histoire évolutive des populations, et de fournir d'importantes informations sur le plan de la génomique fonctionnelle, identifiant notamment des régions responsables de l'expression de traits de caractères (Nielsen, 2005).

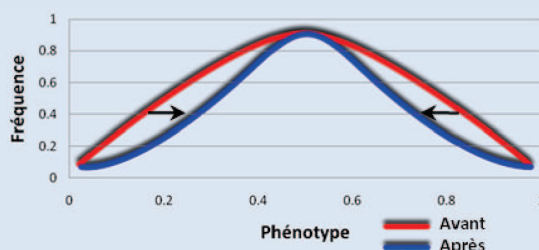
Pour apprécier pleinement le rôle de la génétique dans la sélection, il est nécessaire de considérer qu'il existe différents modèles de sélection (voir **Encadré IV.1**) : la **sélection positive** ; la **sélection purificatrice** ; et la **sélection balancée** (Hurst, 2009). De plus il convient d'introduire les notions de déséquilibre de liaison (« linkage disequilibrium »), d'autostop génétique (« hitch-hiking effect ») et de balayage sélectif (« selective sweep »). Ainsi, bien que l'hypothèse la plus simple soit que chaque locus dans un génome est indépendant, ce modèle n'est pas réaliste. En outre un lien peut entraîner la non-indépendance entre les allèles, ce concept est qualifié de déséquilibre de liaison (DL). Le déséquilibre de liaison peut permettre ainsi à des allèles non bénéfiques (voire délétères) de se propager par "autostop génétique" et pose donc des limites à la sélection (Hurst, 2009).

La sélection positive d'un allèle avantageux entraîne par autostop génétique la fixation dans la population des polymorphismes neutres avoisinant et une distribution de fréquences alléliques biaisée (Maynard-Smith & Haigh, 1974; Fay & Wu, 2000). On appelle balayage sélectif, la réduction de diversité consécutive à la fixation de la mutation favorable dans la population. L'étendue du balayage sélectif va dépendre de l'intensité de la sélection et du taux de recombinaison local (Maynard-Smith & Haigh, 1974; Kaplan *et al.*, 1989; Stephan *et al.*, 1992). Un balayage sélectif complet et récent se traduit par une réduction locale de diversité et l'absence de déséquilibre de liaison (Kim & Stephan, 2000, 2002; Przeworski, 2002; McVean, 2007).

Encadré IV.1. Les modes de sélections (Hurst, 2009)

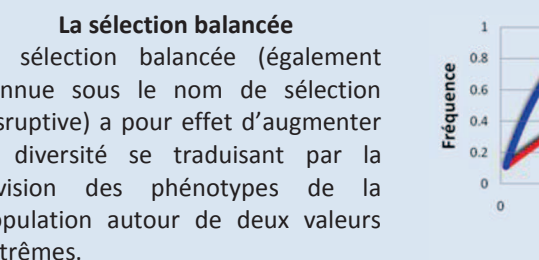
La sélection positive

La sélection positive (également connu sous le nom de sélection darwinienne ou de la sélection directionnelle) entraîne la fixation d'un allèle qui augmente la « fitness » des individus, et se traduit par un déplacement des phénotypes de la population vers un optimum.



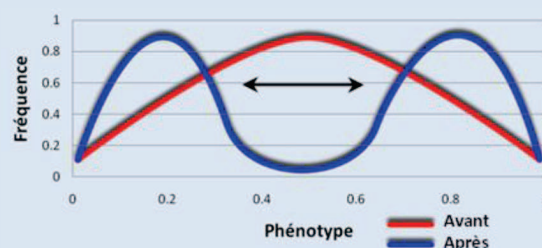
La sélection purificatrice

La sélection purificatrice (également connue sous le nom de sélection négative ou stabilisante) a pour effet d'éliminer les mutations délétères, et se traduit par une diminution de la variabilité phénotypique.



La sélection balancée

La sélection balancée (également connue sous le nom de sélection disruptive) a pour effet d'augmenter la diversité se traduisant par la division des phénotypes de la population autour de deux valeurs extrêmes.



Avec l'avènement de la biologie moléculaire, de nombreux tests statistiques ont été développés pour mettre en évidence les signatures de la sélection (Nielsen, 2001; Storz, 2005; Siol *et al.*, 2010). L'hypothèse de base des procédures statistiques permettant de mettre en évidence les traces de sélection est que la grande majorité des polymorphismes en ségrégation soit neutre. La recherche d'événements sélectifs se fait ainsi en comparant

les données de séquences observées aux patrons de diversité attendus sous cette hypothèse de neutralité de Kimura (Kimura, 1985) et peuvent se regrouper en cinq groupes majeurs présentés dans l'**Encadré IV.2** et détaillés ci-après.

Les tests associés à la variation des niveaux de polymorphisme des séquences constituent une première catégorie de tests. Ils sont très couramment utilisés avec, en particulier, les estimateurs π (Nei, 1987) et θ_w (Watterson, 1975). Une telle approche a été notamment utilisée avec succès chez la Drosophile (*Drosophila melanogaster*) (Beisswanger & Stephan, 2008) ou le tournesol (*Helianthus annuus*) (Kane & Rieseberg, 2008). On notera cependant que ces tests sont relativement sensibles à l'histoire démographique des populations étudiées.

Encadré IV.2. Les tests de sélection (Siol *et al.*, 2010)

Catégorie des tests basés sur :	Signature détectée	Limites du test
Le niveau de diversité : π (Nei, 1987), θ_w (Watterson, 1975), HKA (Hudson <i>et al.</i> , 1987)...	Écart significatif du polymorphisme observé (plus faible ou plus fort) à l'attendu neutre.	Très sensible à l'histoire démographique des populations
Le spectre des fréquences des sites polymorphes: D (Tajima, 1989), D* et F* (Fu & Li, 1993)	Modification dans les proportions relatives des fréquences de mutations	Très sensible à l'histoire démographique des populations
Le déséquilibre de Liaison (DL)	Une augmentation de la fréquence de longs haplotypes du fait de l'augmentation du DL dans cette région	Très sensible à la structure génétique des populations. Le DL décroît rapidement après un balayage sélectif complet
Le ratio des mutations synonymes/non-synonymes : K_A/K_S , M-K test (McDonald & Kreitman, 1991)...	Les différences et les divergences du ratio K_A/K_S entre les loci/populations	Ne permet pas de distinguer des sélections récentes ou passées. Très sensible à la structure génétique des populations.
La différenciation génétique des populations : F_{ST} (Hudson <i>et al.</i> , 1992) G_{ST} (Nei, 1973)...	Une augmentation ou une diminution de la différenciation des populations relative à une région génomique donnée par rapport au génome global.	Très sensible aux sous-structures hiérarchiques des populations (Excoffier <i>et al.</i> , 2009). Nécessite un large échantillon.



Chapitre IV : Étude du polymorphisme des gènes candidats

Les tests basés sur la variation du spectre des fréquences des mutations sont également très usités. Si ces tests dits de neutralité sélective (test de Tajima, 1989 et tests de Fu et Li, 1993) permettent de mettre en évidence des modifications significatives des fréquences de mutations, en revanche, ils sont également sensibles aux effets démographiques entraînant un risque non négligeable de faux-positifs (Siol *et al.*, 2010). Ce type de tests a permis, par l'approche « gène candidat », d'identifier la force et le type de sélection agissant sur des gènes particuliers. L'approche « gènes candidats » est une démarche avec « a priori », dite « top-down ». L'idée est de partir de l'observation d'une variation phénotypique pour étudier le gène ou le groupe de gènes impliqués dans le processus physiologique observé. Cette stratégie nécessite donc de connaître parfaitement les déterminants génétiques de la variation phénotypique observée, ce qui constitue une limite évidente à ce type d'approche. Cependant, la stratégie gène candidat associée à des tests de neutralité sélective a été appliquée avec succès sur plusieurs espèces de plantes notamment pour des gènes liés à la tolérance à la sécheresse (Kane & Rieseberg, 2007; Giombini *et al.*, 2009; Philippe *et al.*, 2010; Xia *et al.*, 2010) ou à la tolérance au froid (Zhen & Ungerer, 2008; Eckert *et al.*, 2009; Wachowiak *et al.*, 2009).

De nombreux tests ont été développés à partir de l'observation de variations du déséquilibre de liaison (Hudson *et al.*, 1994; Kelly, 1997; Depaulis & Veuille, 1998; Andolfatto *et al.*, 1999; Sabeti *et al.*, 2002; Kim & Nielsen, 2004). Cependant, ces tests sont sensibles à la structure génétique des populations étudiées (Siol *et al.*, 2010). La mise en évidence d'une sélection positive par signature de déséquilibre de liaison laissée par un balayage sélectif, n'est possible que durant un court laps de temps car le DL tend à disparaître après que la mutation sélectionnée ait été fixée dans la population (Przeworski, 2002; McVean, 2007).

Les tests basés sur le ratio des mutations synonymes/non-synonymes se basent sur un postulat très simple. En absence de sélection, les substitutions synonymes et non-synonymes ont une fréquence identique et le ratio $K_A/K_S = 1$ (Yang & Bielawski, 2000; Siol *et al.*, 2010). Cependant ce type de test ne permet pas de distinguer les mutations récentes des mutations anciennes et se révèle très sensible à la structure des populations.

Enfin, une dernière catégorie de tests couramment utilisés se base sur l'analyse de la différenciation entre les populations, relative à un locus donné. La différenciation à ce locus

est alors comparée à celle observée sur des loci neutres afin de détecter des écarts significatifs. L'idée ici est très simple, si la sélection a favorisé différents allèles dans des populations distinctes, une augmentation de différenciation F_{ST} est attendue entre les populations au locus sélectionné, en comparaison à l'attendu neutre (Charlesworth & Eyre-Walker, 2006). Si la sélection a favorisé le même allèle dans différentes populations alors la différenciation observée au locus est plus faible comparée à celle attendue sous une simple dérive génétique. Le principal problème que soulève cette approche est de déterminer avec certitude la distribution attendue des F_{ST} sous l'hypothèse nulle de neutralité sélective (Sjol *et al.*, 2010). Dans une étude récente, Excoffier démontre notamment que des sous-structures hiérarchiques des populations induisent la détection de faux-positifs par ce type d'approche (Excoffier *et al.*, 2009).

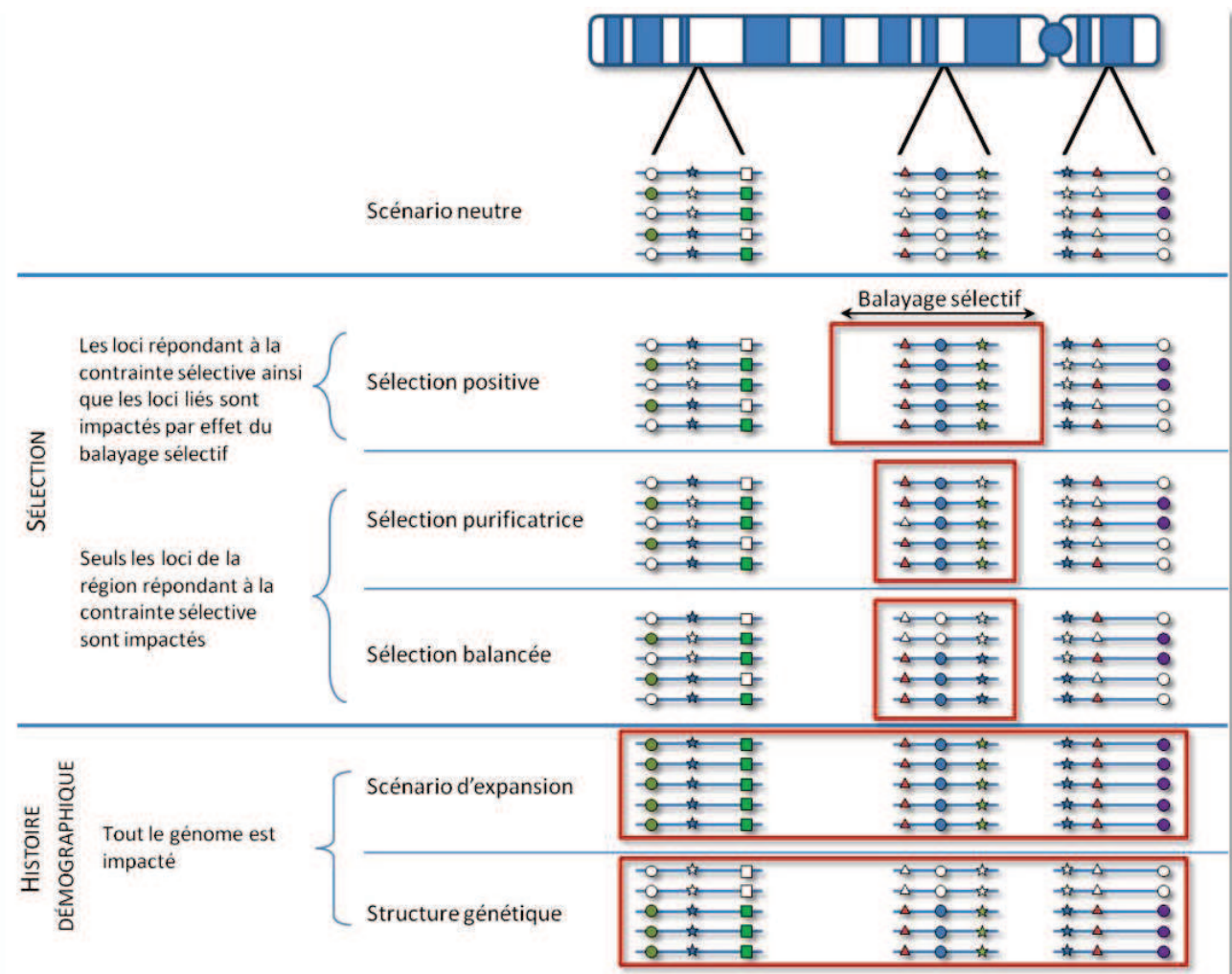


Figure IV.1. Impact attendu de la sélection naturelle et des phénomènes démographiques sur 9 loci de trois régions différentes d'un chromosome. Figure adaptée à partir de (Schlotterer, 2003).



Afin d'illustrer l'impact des différents types de sélection et les risques de faux positifs liés aux processus démographiques, nous proposons un exemple adapté de l'étude de Schlotterer (Schlotterer, 2003) à la **Figure IV.1**. La figure montre la diversité observée au sein de neuf loci provenant de trois régions différentes d'un chromosome. En absence de sélection, les neuf loci évoluent indépendamment (**Figure IV.1**). Dans les scénarios de sélection purificatrice, (consistant en l'élimination de mutations délétères) ou de sélection positive (conduisant à la fixation rapide d'un allèle favorable) on observe une réduction de la diversité. Lors d'un événement de sélection positive, l'autostop génétique, lié au balayage sélectif, entraîne en outre une diminution drastique de diversité au voisinage du marqueur mis en jeu dans la réponse adaptative (**Figure IV.1**). La sélection balancée est elle responsable du maintien de plusieurs allèles sur le long terme à un locus donné (**Figure IV.1**). Ce type de patron de sélection s'observe en particulier lors de phénomènes d'adaptation locale d'un groupe d'individus d'une population. Comme nous venons de le voir, les risques de confondre des traces laissées par la sélection et celles dues à des événements démographiques sont importants. La **Figure IV.1** montre notamment qu'un événement d'expansion d'une population entraînera le même patron de diversité qu'une sélection positive au locus considéré. De même, une forte structure génétique peut mimer un patron de sélection disruptive (**Figure IV.1**). Cependant, contrairement aux patrons de sélection, les phénomènes démographiques ne ciblent pas des loci en particuliers mais affectent tout le génome de la même façon (Cavalli-Sforza, 1966) (**Figure IV.1**).

1.2 Bases moléculaires de la variation des Acides Gras (SAD)

Nous avons mis en évidence chez *Vitellaria paradoxa*, une variation pour les teneurs relatives en acides oléique et stéarique (cf. Chapitre III). Chez les plantes (**Figure IV.2**), la formation d'acide oléique C18 : 1 est catalysée par une enzyme soluble du chloroplaste : la stéaroyl-acyl-carrier-protéine-désaturase (S-ACP-DES) (Somerville *et al.*, 2000). Cette enzyme catalyse spécifiquement la conversion de l'acide stéarique (C18 : 0) en acide oléique (C18 : 1) et son activité est fortement responsable des ratios entre les acides gras saturés et insaturés (Somerville *et al.*, 2000). Ainsi, dans notre étude, cette protéine apparaît comme un candidat majeur dans l'explication des patrons de variation observés entre l'Est et l'Ouest de l'aire naturelle du Karité.

Chapitre IV : Étude du polymorphisme des gènes candidats

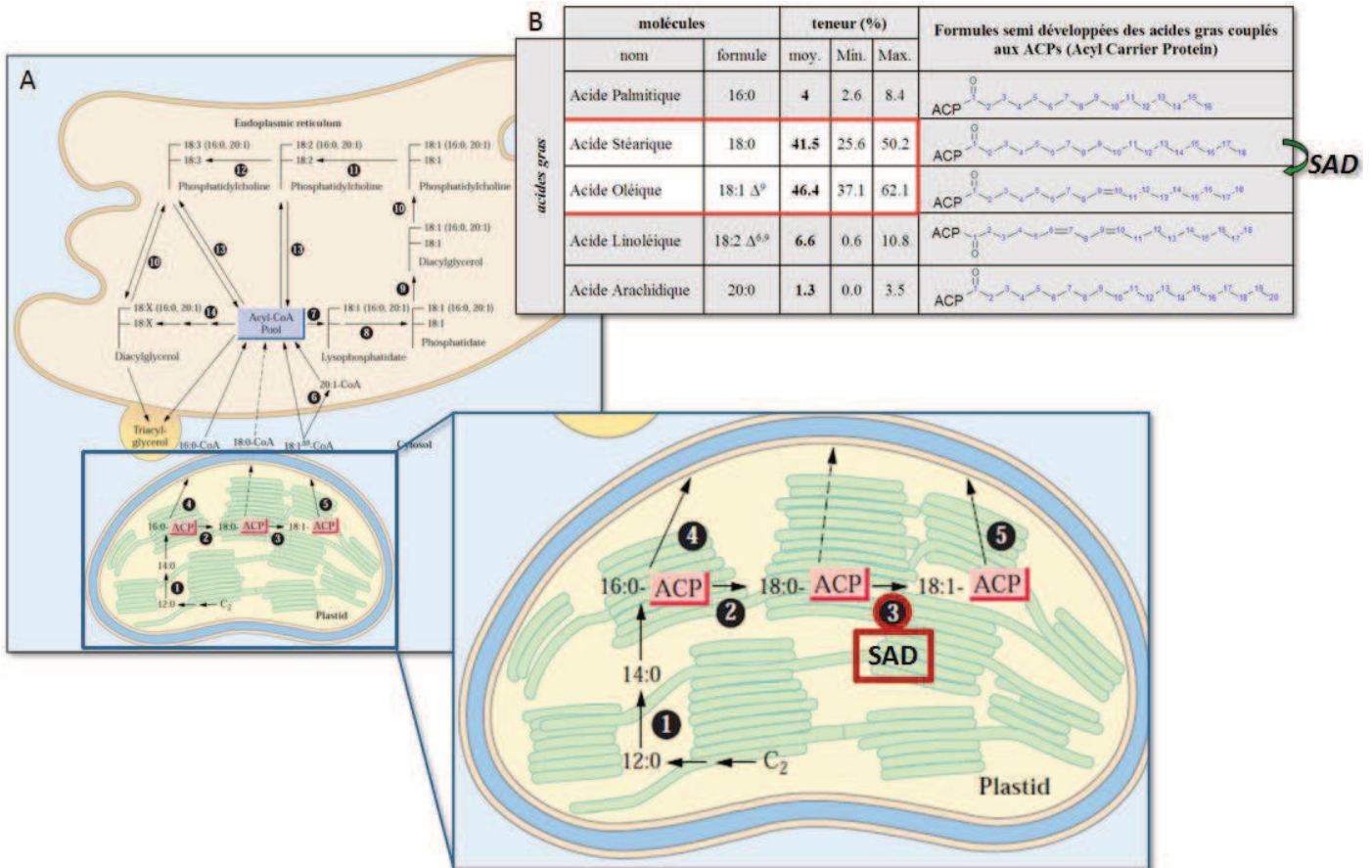


Figure IV.2. Voie de Biosynthèse des acides gras. L'encadré à droite représente un zoom de la partie inférieure de la voie de biosynthèse. Les flèches noires représentent les différentes étapes de la voie de biosynthèse. Les enzymes impliquées à chaque étape sont indiquées dans les cercles noirs. (1) Fatty acid synthase 1 (FAS1) ; (2) Fatty acid synthase 2 (FAS2) ; (3) stéaroyl-ACP désaturase (SAD) ; (4) palmitoyl-ACP thioestérase ; (5) oléoyl-ACP thioestérase ; (6) oléate élongase ; (7) acyl-CoA: glycérol-3-phosphate acyltransférase ; (8) acyl-CoA:lysophosphatidate acyltransférase ; (9) phosphatidate phosphatase ; (10) CDP-choline: diacylglycérol cholinephosphotransférase ; (11) oléate désaturase ; (12) linoleate desaturase ; (13) acyl-CoA : sn-1 acyllysophosphatidylcholine acyltransférase ; (14) identiques à (7), (8), et (9).

Cette enzyme (S-ACP-DES) bien connue de la voie de biosynthèse des acides gras (**Figure IV.2**) est active à l'état homo-dimérique (**Figure IV.3**). Elle a en outre été purifiée à partir de plusieurs plantes et les gènes codant pour la sous-unité protéique ont été caractérisés chez de nombreuses espèces (Shanklin *et al.*, 1991; Shanklin & Somerville, 1991; Thompson *et al.*, 1991; Cahoon *et al.*, 1994; Cahoon *et al.*, 1996; Cahoon *et al.*, 1998; Whittle *et al.*, 2005). La structure tridimensionnelle et cristallographique de la protéine a été déterminée chez le ricin (*Ricinus communis*) (**Figure IV.3**), mettant en évidence que chaque monomère possède un centre catalytique di-ferrique positionné derrière un faisceau de quatre alpha-hélices (Schneider *et al.*, 1992; Lindqvist *et al.*, 1996; Moche *et al.*, 2003). Ainsi, de part la fonction majeure de la S-ACP-DES dans la composition des plantes en acide gras, de nombreuses études se sont intéressées à cette enzyme dans des stratégies

d'amélioration qualitative ou quantitative des productions d'huiles végétales (Knutzon *et al.*, 1991; Knutzon *et al.*, 1992; Fofana *et al.*, 2006; Pleite *et al.*, 2006; Mar-Heyming *et al.*, 2008; Venegas-Caleron *et al.*, 2009). Les résultats obtenus au cours de ces études ont notamment permis de vérifier l'importance de la protéine dans le contrôle des proportions relatives en acide gras saturés et insaturés.

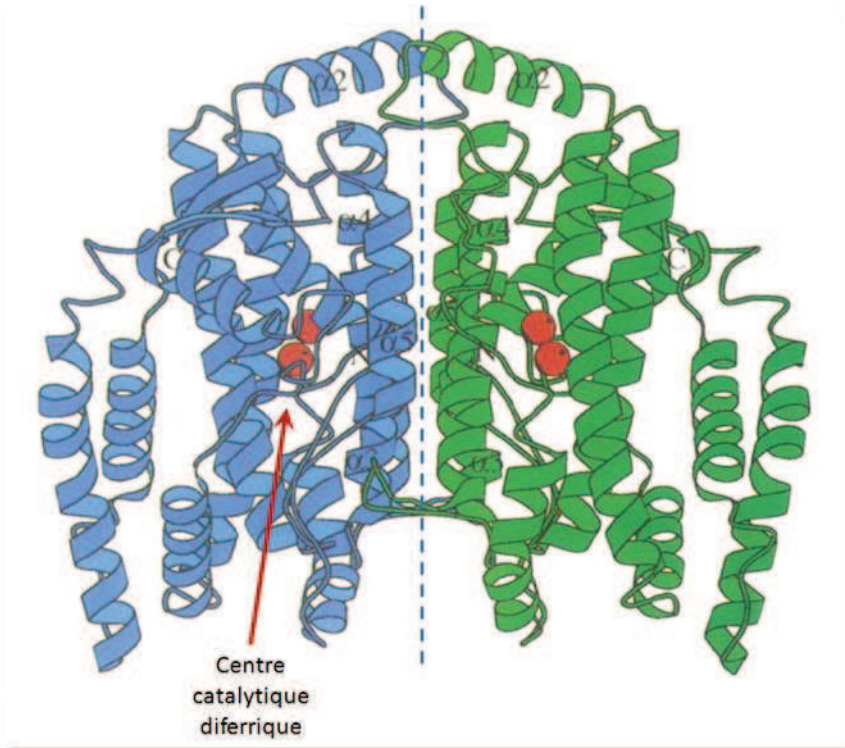


Figure IV.3. Vue tridimensionnelle du dimère protéique de S-ACP-DES chez *Ricinu communis* (Lindqvist *et al.*, 1996)

L'exploration récente du génome d'*Arabidopsis thaliana* révèle la présence de sept gènes S-ACP-DES-LIKE (Kachroo *et al.*, 2007). On parle alors de S-ACP-DES comme une famille de gènes comprenant divers isoformes. Parmi les sept isoformes d'*Arabidopsis*, le gène *SAD* (nommé dans la littérature *SSI2* ou *FAB2*) tient cependant une place prépondérante pour la désaturation de l'acide stéarique en acide oléique (Cao *et al.*, 2010). La littérature montre que l'extinction de ce gène *SAD* induit des phénotypes très perturbés chez *Arabidopsis thaliana* (Figure IV.4). Les mutants *sad* (*ssi2/fab2*) arborent ainsi des taux accrus en acide stéarique et réduits en acide oléique (Lightner *et al.*, 1994; Kachroo *et al.*, 2001); ils présentent un phénotype nain (Kachroo *et al.*, 2001; Shah *et al.*, 2001); et des mécanismes de défense déficientes (Kachroo *et al.*, 2001; Kachroo *et al.*, 2003a; Kachroo *et al.*, 2003b; Kachroo *et al.*, 2004; Kachroo *et al.*, 2005). À l'inverse, l'altération des autres isoformes n'impacte que très peu le phénotype d'*Arabidopsis thaliana*, n'influe pas la quantité d'acide

oléique et n'induit pas de défaut de signalisation du système moléculaire de défense (Kachroo *et al.*, 2007; Kachroo *et al.*, 2008). En outre, les six isoformes autres que *SAD* ne permettent pas de palier le déficit en acide oléique des mutants *sad*.



Figure IV.4. Comparaison des morphologies du type sauvage SAD (SSI2) (à gauche) et du type mutant *ssi2* (à droite) chez *Arabidopsis thaliana* (Kachroo *et al.*, 2003a)

1.3 Bases moléculaires de la variation des tocophérols (*VTE2-2*)

Nous avons vu dans le Chapitre précédent qu'il existe une grande variabilité des teneurs en tocophérols au sein et entre les populations de *Vitellaria paradoxa*. Ces antioxydants amphiphiles très puissants constituent notamment chez les organismes photosynthétiques des acteurs importants du système de défense contre la peroxydation des composés lipidiques et les formes réactives de l'oxygène (Kamal-Eldin & Appelqvist, 1996; Abbasi *et al.*, 2007; Collin *et al.*, 2008; Brigelius-Flohé, 2009; Falk & Munné-Bosch, 2010; Mène-Saffrané & Dellapenna, 2010). Les niveaux de tocophérols sont connus pour changer de façon significative au cours de la croissance et le développement des plantes, ainsi qu'en réponse au stress, probablement en conséquence de la modification du niveau d'expression des gènes de la voie de biosynthèse à laquelle ils sont associés (Collakova & Dellapenna, 2003b) (Figure IV.6).



mutant *vte2-1*

sauvage

mutant *vte2-2*

Figure IV.5. Comparaison des morphologies du type sauvage (au centre) et des mutant *vte2-1* (à gauche) et *vte2-2* (à droite) chez *Arabidopsis thaliana* (Maeda *et al.*, 2006)

Comme nous l'avons vu au cours du Chapitre III, la voie de biosynthèse des tocophérols comporte trois étapes préalables avant l'obtention du 2-méthyl-6-phytyl-1,4-benzoquinol, molécule qui devient ensuite le substrat à la synthèse des tocophérols (cf **Figure III.4**). Divers enzymes interviennent dans cette voie de biosynthèse: la *p*-hydroxy-phényl pyruvate dioxygénase (HPPD) (Garcia *et al.*, 1997; Norris *et al.*, 1998; Garcia *et al.*, 1999); la géranylgéranyle diphosphate synthase 1 (GGPS1) (Okada *et al.*, 2000); la géranylgéranyle diphosphate réductase (GGDR) (Addlesee *et al.*, 1996; Keller *et al.*, 1998); une kinase du stroma des chloroplaste (KIN) (Soll *et al.*, 1980); l'homogentisate phytyl transférase (HPT) (Soll, 1987; Collakova & Dellapenna, 2001; Savidge *et al.*, 2002); la 2,3-diméthyl-6-phytyl-1,4-benzoquinol méthyltransférase (MPBQ MT) (Marshall *et al.*, 1985; Soll, 1987; Shintani *et al.*, 2002); la tocophérol cyclase (TC) (Stocker *et al.*, 1996; Arango & Heise, 1998; Porfirova *et al.*, 2002); et la tocophérol-méthyltransférase (D'harlingue & Camara, 1985; Shintani & Dellapenna, 1998) (**Figure IV.6**). Parmi ces enzymes, l'homogentisate phytyl transférase (HPT), enzyme liée à la membrane du chloroplaste, catalyse l'étape de condensation de l'acide homogentisique et du phytyl diphosphate (PDP), pour former le 2-méthyl-6-phytyl-1,4-benzoquinol, première molécule antioxydante et substrat indispensable à la formation des différentes formes de tocophérols : α , β , γ et ζ (**Figure IV.6**). La HPT apparait comme particulièrement limitante dans la synthèse des tocophérols (Collakova & Dellapenna, 2003a; Mène-Saffrané & Dellapenna, 2010). Deux paralogues du gène *VITAMIN E2* (*VTE2* ou *HPT*) codant pour l'enzyme HPT ont été découverts chez *Arabidopsis thaliana* (Sattler *et al.*, 2004; Venkatesh *et al.*, 2006). La surexpression constitutive du gène *VTE2* sous le contrôle d'un promoteur 35S issu du virus de la mosaïque du chou-fleur (CaMV 35S), montre chez les lignées transformantes

The diagram illustrates the biosynthetic pathways for tocopherols (vitamin E) in *Arabidopsis thaliana*, showing the conversion of tyrosine and the DXP pathway into various tocopherol forms.

Tyrosine Pathway:

- Tyrosine is converted to *p*-hydroxyphenyl pyruvate by the enzyme **TAT**.
- p*-hydroxyphenyl pyruvate is converted to homogentisate by the enzyme **HPPD**, releasing CO_2 .
- Homogentisate is converted to PDP (2-methyl-6-phytyl-BQ) by the enzyme **HGAD**, releasing maleylacetoacetate.
- PDP is converted to 2-methyl-6-phytyl-BQ by the enzyme **HPT**, releasing $\text{CO}_2 + \text{PP}_i$.

DXP Pathway:

- 3 isopentenyl-PP and dimethylallyl-PP are converted to GGDP by the enzyme **GGPS1**.
- GGDP is converted to PDP (2-methyl-6-phytyl-BQ) by the enzyme **GGDR**, releasing chlorophyll.
- PDP is converted to 2-methyl-6-phytyl-BQ by the enzyme **KIN**, using ATP.

Tocopherol Biosynthesis:

- 2-methyl-6-phytyl-BQ is converted to 2,3-dimethyl-6-phytyl-BQ by the enzyme **MT**, using SAM.
- 2,3-dimethyl-6-phytyl-BQ is converted to δ -tocopherol by the enzyme **TC**.
- δ -tocopherol is converted to γ -tocopherol by the enzyme **γ -TMT**, using SAM.
- γ -tocopherol is converted to α -tocopherol by the enzyme **γ -TMT**, using SAM.

L'enzyme HTP apparaît donc comme clef dans la voie de biosynthèse des tocophérols. Dans cette étude, nous nous intéresserons à la variabilité nucléotidique de *VTE2-2*, afin de mieux comprendre l'impact des forces évolutives sur la diversité génétique de ce gène et tenter de lier cette variabilité à une variation des teneurs en α -tocophérol et γ -tocophérol au sein d'un gradient climatique Nord-Sud au Mali. Le choix de ce gène est également motivé par la conservation de la structure et des séquences exoniques du gène



VTE2-2 chez les plantes (Collakova & Dellapenna, 2001; Goffman & Becker, 2001b, a; Giovannoni, 2004; Gilliland *et al.*, 2006), nous permettant d'envisager le clonage de ce gène chez *Vitellaria paradoxa*.

1.4 Objectifs et stratégie mise en place dans ce chapitre

La littérature suggère que le rôle adaptatif des constituants chimiques des graines est très important chez les plantes (voir Chapitre III), en particulier concernant les lipides de réserve (Linder, 2000). Dans le chapitre précédent, nous mettons en évidence une forte structuration Est/Ouest de la variabilité chimique des graines de Karité sur son aire naturelle pour la composition en acides gras et tocophérols de la matière grasse. Dans ce contexte, l'objectif général de ce chapitre de la thèse est de mieux comprendre l'impact des facteurs évolutifs sur le polymorphisme nucléotidique de gènes impliqués dans les voies de biosynthèse des acides gras (**Figure IV.2**) et des tocophérols (**Figure IV.6**), et supposés candidats dans le contrôle de la variation phénotypique observée (**Chapitre III**). Dans cet objectif général, nous cherchons à répondre à certaines questions plus spécifiques :

- (1) Comment s'organise la diversité génétique des gènes candidats au sein de l'aire naturelle, où s'expriment des facteurs démographiques et historiques importants ?

Pour cela, nous étudions la structure, le polymorphisme et le déséquilibre de liaison au sein des gènes *SAD* et *VTE2-2* et nous recherchons les indices potentiels de traces de sélection.

- (2) Quel est l'impact de la variation nucléotidique du gène *VTE2-2* sur la variation chimique observée au sein d'un gradient Nord-Sud au Mali ?

Afin de répondre à cette question, nous proposons dans ce chapitre une démarche d'étude du polymorphisme, du déséquilibre de liaison, des traces de sélections, suivie d'une approche par test d'association génotype/phénotype.

2 MATÉRIEL & MÉTHODES

2.1 Échantillonnage

Tableau IV.1. Échantillonnage de l'aire naturelle utilisé pour l'analyse du polymorphisme des gènes *VTE2-2*, *SAD1* et *SA2*. Lat = latitude moyenne du site ; Lon = longitude moyenne du site ; N ind = nombre d'individus échantillonnés ; N seq = nombre de séquences (allèles) utilisées (e.g. 2 maximum par individus).

Échantillon	Pays	Site	Sous-espèce	Lat	Lon	N ind	N seq VTE2-2	N seq SAD1_5'	N seq SAD1_3'	N seq SAD2_5'	N seq SAD2_3'
Aire Naturelle						57	88	32	52	31	52
Afrique de l'Ouest	Sénégal	Bandafassi	<i>ssp. paradoxa</i>	12.53	-12.32	2	4	2	2	1	2
		Samecoutha	<i>ssp. paradoxa</i>	12.6	-12.13	2		2	2	2	2
	Mali	Kodougouni	<i>ssp. paradoxa</i>	12.97	-6.66	1	2				
		Mahina	<i>ssp. paradoxa</i>	13.78	-10.84	2	2	2	2	2	2
		Mperesso	<i>ssp. paradoxa</i>	12.28	-5.33	1		2	2	2	2
		Tennedeli	<i>ssp. paradoxa</i>	14.09	-3.45	1	2				
		Torodo	<i>ssp. paradoxa</i>	14.5	-8.84	1	2				
	Burkina-Faso	Bonogo	<i>ssp. paradoxa</i>	12.06	-1.6	2	2	2	2	2	2
		Fada N'Gourma	<i>ssp. paradoxa</i>	12.06	0.41	1	2				
		Karo	<i>ssp. paradoxa</i>	12.37	-3.3	2	2	2	2	2	2
		Konia	<i>ssp. paradoxa</i>	13.1	-0.99	1	2				
		Peni	<i>ssp. paradoxa</i>	10.97	-4.45	1	2				
		Pô	<i>ssp. paradoxa</i>	11.18	-1.15	1	2				
	Côte d'Ivoire	Korhogo	<i>ssp. paradoxa</i>	9.62	-4.39	2	4		4		4
	Ghana	Kadia	<i>ssp. paradoxa</i>	9.9	-0.86	1	2				
		Kawampe	<i>ssp. paradoxa</i>	8.43	-1.56	1	2				
		Larabanga	<i>ssp. paradoxa</i>	9.19	-1.85	1	2	2	2	2	2
		Yendi	<i>ssp. paradoxa</i>	9.45	-0.04	2	2	2	2	2	2
	Togo	Djangbassou	<i>ssp. paradoxa</i>	7.81	1.19	1	2				
		Kante	<i>ssp. paradoxa</i>	9.99	1.01	1	2	2	2	2	2
		Kolina	<i>ssp. paradoxa</i>	9.1	1.14	1		2	2	2	2
		Koudassi	<i>ssp. paradoxa</i>	6.64	0.86	1	2				
		Tabango	<i>ssp. paradoxa</i>	10.73	0.2	1	2				
	Benin	Cocota	<i>ssp. paradoxa</i>	10.41	1.36	1	2				
		Kalale	<i>ssp. paradoxa</i>	10.27	3.36	1	2	2	2	2	2
		Kandi	<i>ssp. paradoxa</i>	11.16	2.95	1	2				
		Kpale	<i>ssp. paradoxa</i>	9.65	1.73	1	2				
		Pira	<i>ssp. paradoxa</i>	8.48	1.78	2	2	2	2	2	2
	Niger	Boumba	<i>ssp. paradoxa</i>	12.41	2.86	2	4				
	Nigeria	Abuja	<i>ssp. paradoxa</i>	9.08	7.42	2	4				
Afrique Centrale	Cameroun	Bangangte	<i>ssp. paradoxa</i>	5.15	10.52	1	2		2		2
		Foumban	<i>ssp. paradoxa</i>	4.73	10.9	1	2				
		Mafa-Kilda	<i>ssp. paradoxa</i>	9.03	13.75	1	2		2		2
	Tchad	Kelo	<i>ssp. paradoxa</i>	9.2	15.78	2	2		2		2
		Mondou	<i>ssp. paradoxa</i>	8.62	16.02	1	2		2		2
		Sarh	<i>ssp. paradoxa</i>	9.17	18.37	1	2		2		2
	Centrafrique	CENTRE	<i>ssp. paradoxa</i>	8.25	20.61	3	4		6		6
		SUD	<i>ssp. paradoxa</i>	7.9	20.85	1	2				
Afrique de l'Est	Ouganda	Adumara	<i>ssp. nilotica</i>	2.53	33.05	2	2	4	4	4	4
		Olwoko	<i>ssp. nilotica</i>	2.43	33.5	1	2				
		Otongere	<i>ssp. nilotica</i>	2.53	33.05	1	2				
		Otuke ss	<i>ssp. nilotica</i>	2.43	33.5	2	2	4	4	4	4

Afin de répondre aux objectifs spécifiques de ce chapitre, nous avons utilisé deux échantillonnages différents. Ainsi, l'analyse du polymorphisme des gènes *SAD1*, *SAD2* et *VTE2-2* a été menée sur un premier échantillon représentatif de l'aire naturelle de *Vitellaria paradoxa* (**Tableau IV.1**). Cet échantillon se constitue de 57 individus, répartis dans 42 sites, originaires de 13 pays d'Afrique sub-sahélienne et comportant les deux sous-espèces décrites, *Vitellaria paradoxa* ssp. *paradoxa* et *Vitellaria paradoxa* ssp. *nilotica*. Pour chaque portion de gène étudiée, un nombre variable d'individus a pu être séquencé pour chacun de ses deux allèles (**Tableau IV.1**). Un second échantillonnage, restreint au Mali, a été effectué selon un gradient climatique Nord-Sud. Les individus de ce gradient correspondent à ceux dont les fruits ont été caractérisés chimiquement pour les profils d'acides gras et les teneurs en tocophérols (voir Chapitre III). Au sein de cet échantillonnage nous avons limité l'analyse au gène *VTE2-2*, et avons pu obtenir les deux allèles de 38 individus en provenances de 4 sites, avec du Sud au Nord : Nafégué, Mperesso, Daelan et Sassambourou (cf **Figure III.5**). Le **Tableau IV.2** présente l'échantillon Nord-Sud du Mali.

Tableau IV.2. Échantillonnage du gradient Nord-Sud au Mali utilisé pour l'analyse du polymorphisme des gènes *VTE2-2*, et les tests d'association phénotype/génotype (*VTE2-2*/tocophérols). Lat = latitude moyenne du site ; Lon = longitude moyenne du site ; N ind = nombre d'individus échantillonnés ; N seq = nombre de séquences (allèles) obtenus.

Échantillon	Pays	Site	Sous-espèce	Lat	Lon	N ind	N seq <i>VTE2-2</i>
Gradient Nord-Sud Mali						38	76
	Mali	Nafégué	ssp. <i>paradoxa</i>	10.51	-5.98	10	20
		Mperesso	ssp. <i>paradoxa</i>	12.33	-5.6	11	22
		Daelan	ssp. <i>paradoxa</i>	13.25	-4.99	10	20
		Sassambourou	ssp. <i>paradoxa</i>	14.31	-3.5	7	14

2.2 Méthodes de laboratoire

2.2.1 Comptage chromosomique

Afin d'étudier le niveau de ploïdie des deux sous-espèces de *Vitellaria paradoxa*, nous avons prélevé les apex racinaires de 8 individus de type *paradoxa* et de 2 individus de type *nilotica*. Les apex racinaires sont particulièrement intéressants pour le protocole car ils présentent une forte activité mitotique. Une fois prélevés, les apex sont plongés dans 1 ml de solution à 0,03% 8-hydroxy-quinoléine à 20-22 ° C pour (4-9) h. Les apex sont ensuite transférés dans une solution fixation pendant environ 12 h (nuit). Les apex sont ensuite



placés et pressés (avec suffisamment de force pour répandre les chromosomes hors des cellules) entre lame et lamelle avec une goutte de solution de coloration. Les lames obtenues sont placées sur la plate-forme d'un microscope optique à grossissement 200-400X. Les images sont ensuite photographiées et numérisées par un appareil photographique haute définition relié à un ordinateur.

2.2.2 Extraction de l'ADN génomique

L'ADN des individus sélectionnés a été extrait par la méthode « dite » de Dolezel (Ky *et al.*, 2000). L'ADN a été quantifié sur gel d'agarose. Pour cela nous avons déposé 1 μ L d'ADN ainsi qu'une gamme de masse connue, de 10 μ g à 100 μ g d'ADN de phage Lambda, sur un gel à 1% d'agarose dans du tampon TAE 1X. Après migration, les concentrations en ADN de chaque individu ont été déterminées par analyse d'image grâce au logiciel IMAGEJ (<http://rsb.info.nih.gov/ij/index.html>). Les concentrations en ADN sont ensuite ajustées afin d'obtenir une concentration finale de 20 ng/ μ L adaptée pour réaliser les PCR.

2.2.3 Extraction de l'ARN messager

L'extraction de l'ARN messager (ARNm) a été effectuée selon une adaptation de la méthode développée par Zhang (Zhang *et al.*, 2006), incluant différentes étapes de précipitation au chlorure de lithium et des ultra-centrifugations (30.000 g). Une fois extrait, l'ARNm est mis en solution dans une eau « RNAase-free » puis a été dosé sur gel d'agarose et en utilisant un NANODROP (NanoDrop Technologies, Wilmington, DE, <http://www.nanodrop.com>).

2.2.4 RACE (Rapid Amplification of cDNA Ends)

Afin de cloner les séquences des gènes *SAD* et *VTE2-2*, nous avons utilisé une méthodologie de type RACE (Rapid Amplification of cDNA Ends) (Frohman *et al.*, 1988) (**Annexe IV.1**). Cette approche a été permise par l'amplification et le séquençage au préalable d'un petit fragment de chaque gène étudié par l'intermédiaire d'amorces dégénérées définies dans les zones conservées des gènes (les zones conservées ont été caractérisées par l'alignement global de multiples séquences de ces gènes chez d'autres espèces de plantes récupérées dans les bases de données EMBL/NCBI). Avant la mise en

place de la RACE, l'ADN complémentaire (ADNc) premier brin est obtenu par reverse transcription de l'ARN total ou messenger, à l'aide du kit RevertAid™ H Minus Strand cDNA Synthesis kit.

Pour la RACE 5', une queue poly-C est créée à l'extrémité 3' du cDNA 1^{er} brin grâce à la TDTase (Terminal Desoxynucleotide Transferase). Cette queue sert de second point d'ancrage pour l'amplification de région 5' adjacente à la région connue, tandis que pour le RACE 3' on utilise la présence de la queue poly-A à l'extrémité 3' de l'ARNm. Une série de PCR nichée est effectuée afin d'amplifier spécifiquement les extrémités 3' et 5' de l'ADNc des gènes candidats. Les conditions d'amplification PCR sont les suivantes : 5 µL Tampon PCR 10X, 1.5mM final de MgCl₂, 0,4 µM final d'amorces sens et antisens à 10 µM, 0,2 unités de Taq DNA polymérase et 5 µL de produit de reverse transcription contenant l'ADNc. Le programme PCR a été effectué sur un thermocycler Mastercycler Eppendorf® et comprend un touch-down pour maximiser la spécificité de l'amplification tout en optimisant son rendement.

2.2.5 Amplification des gènes candidats par PCR

Les séquences d'ADNc obtenues par les techniques de RACE nous permettent de définir de nouvelles amorces spécifiques pour l'amplification des gènes ciblés sur l'ADN génomique (ADNg) des individus de *Vitellaria paradoxa*. Les PCR ont été effectuées dans un volume total de 10 µL selon le protocole suivant : 6,58 µL d'H₂O, 1 µL de tampon 10X, 0,3 µL d'amorce U à 10 µM, 0,3 µL d'amorce L à 10 µM, 0,08 µL de dNTP à 25 mM, 0,5 µL de DMSO (5% final), 0,24 µL de Taq polymérase et 1 µL d'ADN à 20 ng/µL. L'amplification a été réalisée à l'aide d'un thermocycleur Eppendorf®.

2.2.6 Clonage des produits PCR

Cette étape est effectuée par le kit pGEM-T Vector System I de Promega® conformément aux préconisations du fabricant. L'incubation se déroule toute la nuit à 4°C. Les milieux de cultures LB agar sont préparés puis autoclavés le jour même de la transformation bactérienne. La concentration finale en ampicilline est de 0,1 mg/mL; en X-gal de 80 µg/mL; et en IPTG de 20 mM. Pour la transformation bactérienne, nous avons utilisé les bactéries XL10-Gold ultracompetent cells. Une heure avant utilisation, les

bactéries ultra-compétentes sont transférées de -80°C à -20°C . Après ajout de $2\ \mu\text{L}$ de β -mercaptoéthanol, une incubation de 10 minutes à 0°C est effectuée. On ajoute ensuite $2\ \mu\text{L}$ de produit de ligation, puis après une nouvelle incubation de 30 minutes à 0°C , on applique un choc thermique de 30 secondes à 42°C . On ajoute alors $450\ \mu\text{L}$ d'un milieu SOC ultra-riche, et les bactéries sont ensuite mises en incubation à 37°C pendant une heure en agitation. On étale ensuite, sur les boîtes de pétries contenant le milieu de culture, 50 à $200\ \mu\text{L}$ de suspension bactérienne. Les boîtes de pétries sont enfin mises en incubation toute la nuit à 37°C . Seules les bactéries transformées par un plasmide vont survivre sur le milieu contenant l'antibiotique (ampicilline). Le test Blanc/Bleu permet un premier screening des bactéries transformées. Les colonies bactériennes ayant un plasmide non recombinant vont posséder un gène LacZ fonctionnel. Elles vont donc avoir une activité galactosidase, permettant la transformation du X-gal en un produit de couleur bleue. À l'inverse, les bactéries transformées par un plasmide recombinant perdent cette activité galactosidase et sont blanches.

2.2.7 Séquençage

Le séquençage des brins sens et antisens, des clones recombinants (au minimum 12 clones par portion de gène et par individu) et des produits PCR le cas échéant (gène *VTE2-2*), a été effectué par la société GATC-Biotech (Allemagne).

2.3 Méthodes d'analyses statistiques

2.3.1 Traitement, reconstruction des phases haplotypiques et alignement des données de séquences

Pour les fragments *SAD1_5'*, *SAD2_5'*, *SAD1_3'*, *SAD2_3'*, étant passé par une étape de clonage et de transformation bactérienne, les données de séquences brutes (chromatogrammes) ont été éditées à l'aide du logiciel CHROMAS (<http://www.technelysium.com.au/chromas.html>) afin d'éliminer les amorces et les résidus de vecteurs plasmidiques. Pour le gène *VTE2-2*, nous ne sommes pas passés par une étape de clonage. Un travail de reconstruction des haplotypes a donc été nécessaire pour les séquences *VTE2-2*. Pour ce faire nous avons bénéficié d'une part de nombreux individus homozygotes complets (27 individus au total) ou bien ne portant qu'un seul point de

polymorphisme et d'autre part de la présence récurrente d'insertions/délétions (INDEL) dans les séquences. Ces INDEL créent un décalage de lecture dans le chromatogramme, grâce à l'information du séquençage effectué dans les deux sens de lecture, il est alors possible manuellement et sans ambiguïté de reconstruire chacun des deux allèles. Ce travail a cependant été facilité par le très faible polymorphisme SNP (Single Nucléotide Polymorphism) détecté au sein des séquences. Les séquences haplotypiques des différentes portions de gènes étudiées ont ensuite été alignées et assemblées en contig indépendamment à l'aide de l'algorithme MUSCLE du logiciel CODONCODE ALIGNER (v. 3.5.6, Codon Code Corporation). Les séquences codantes pour les gènes *VTE2-2* et *SAD* ont été identifiées par BLASTN et BLASTX à l'aide des informations relatives à ces gènes (ARNm) sur la base de données publique NCBI.

2.3.2 Analyse du polymorphisme

Le polymorphisme des séquences nucléotidiques a été analysé à l'aide du logiciel DNASP v. 5.10 (Librado & Rozas, 2009). Pour chaque portion de gène étudié, nous avons estimé la diversité nucléotidique par : S , le nombre de sites en ségrégation ; Hap, le nombre d'haplotypes ; H_d , la diversité haplotypique (Nei, 1987) et deux mesures de diversité (estimateurs de $4N\mu$) couramment utilisées, π (Nei, 1987) et θ_w (Watterson, 1975). Dans un alignement de séquence donné, π mesure le nombre moyen de différences nucléotidiques entre les paires de séquences d'un échantillon (Tajima, 1983) et θ_w mesure le nombre moyen de sites en ségrégation (Watterson, 1975). Ces deux mesures prennent en compte la taille de l'échantillon (nombre de séquences) et sont en général exprimées par site. Contrairement à π , θ_w ne dépend pas des fréquences alléliques puisqu'il est basé uniquement sur le nombre de sites polymorphes. L'évolution des valeurs de π le long des séquences étudiées a été évaluée par la méthode de « sliding window » (fenêtre coulissante) avec une taille de fenêtre de 50 nucléotides et un pas de déplacement de 5 nucléotides grâce au logiciel DNASP v. 5.10 (Librado & Rozas, 2009). Afin de démontrer si les niveaux de polymorphisme diffèrent de l'attendu neutre apporté sous le modèle de Wright-Fisher, des simulations de coalescence ont été réalisées grâce au logiciel DNASP v. 5.10 (Librado & Rozas, 2009).

2.3.3 Tests de neutralité



Chapitre IV : Étude du polymorphisme des gènes candidats

Nous avons testé les écarts par rapport au modèle standard neutre afin d'identifier des signatures de sélection au sein des différents fragments analysés. Ces analyses ont été réalisées à l'aide du logiciel DNAsP v. 5.10 (Librado & Rozas, 2009) en utilisant plusieurs statistiques: le D de Tajima (Tajima, 1989), et le D* et le F* de Fu et Li (Fu & Li, 1993). Ces statistiques exploitent l'information contenue dans la distribution des fréquences de polymorphismes. Le D de Tajima (Tajima, 1989) permet de détecter un excès ou un déficit de variants rares relativement à l'attendu neutre au sein de la région avoisinant un site soumis à sélection. Il se base pour cela sur la comparaison entre θ et π . Sous l'hypothèse de neutralité, le D de Tajima a pour espérance zéro. Ainsi, on attend sous une sélection balancée le maintien d'un excès de polymorphismes dont les allèles sont en fréquences équilibrées, ce qui se traduit par un D de Tajima positif. Inversement, les événements de sélection directionnelle récents impliquant la sélection de polymorphismes initialement en très faible fréquence, entraînent une réduction drastique de la diversité par l'effet du balayage sélectif. Par la suite, de nouvelles mutations apparaissent, qui n'ont pas encore atteint leur fréquence d'équilibre, ce qui se traduit par un excès de variants rares par rapport à l'attendu neutre et un D de Tajima est négatif. Les régions soumises à sélection purifiante se caractérisent également par un D de Tajima négatif, mais dans une moindre mesure que les événements de sélection positive/directionnelles. Plus spécifiquement, alors que la sélection positive affecte une région dans son ensemble (site sélectionné et sites neutres) du fait de sa rapidité qui limite la survenue d'événements de recombinaison, la sélection purifiante se fait sur une longue durée, ce qui permet de distinguer l'histoire des sites sélectionnés de celle des sites neutres qui ont le temps d'accumuler des mutations.

Les D* et F* de Fu et Li (1993) sont très comparables au D de Tajima mais se basent sur la comparaison entre θ ou π et le nombre de singletons (mutation présente une seule fois dans l'échantillon). Ces statistiques ont une puissance supérieure au D de Tajima pour détecter les événements de sélection directionnelle qui se caractérisent par des mutations récentes, souvent à l'état de singletons (Fu & Li, 1993). Ces dernières statistiques, tout comme le D de Tajima, peuvent être biaisées par l'impact des effets démographiques sur la diversité.

De même que pour les indices de diversité, l'évolution des valeurs de D, D*, et F* le long des fragments de gènes utilisés, a été évaluée par la méthode de « sliding window »



(fenêtre coulissante) avec une taille de fenêtre de 25 nucléotides et un pas de déplacement de 5 nucléotides grâce au logiciel DNASP v. 5.10 (Librado & Rozas, 2009). La significativité des tests de neutralité a été obtenue par des simulations de coalescence DNASP v. 5.10 (Librado & Rozas, 2009).

2.3.4 Différenciation des populations

La structure génétique des populations a été mesurée en calculant l'indice de différenciation F_{ST} (Hudson *et al.*, 1992) et le G_{ST} haplotypique (Nei, 1973) entre les groupes de populations (clusters) définis par l'analyse Bayésienne de la structure des populations dans le Chapitre II de cette thèse (Afrique de l'Ouest/Afrique Centrale/Afrique de l'Est). Le F_{ST} considère indépendamment chaque position nucléotidique et le G_{ST} considère chaque haplotype comme un allèle distinct. En particulier, sous l'hypothèse d'une sélection disruptive (= adaptation locale) agissant sur les gènes candidats, on s'attend à des F_{ST} et G_{ST} forts entre les deux populations. Le F_{ST} et le G_{ST} ont été calculés en utilisant le logiciel DNASP v. 5.10 (Librado & Rozas, 2009). La significativité des indices de différenciation a été obtenue par l'application d'un test d'exactitude de Fisher grâce à DNASP v. 5.10 (Librado & Rozas, 2009).

2.3.5 Constructions des arbres phylogénétiques et des réseaux haplotypiques

Les relations phylogénétiques entre les haplotypes observés ont été obtenues par reconstruction des réseaux haplotypiques grâce au logiciel NETWORK 4.516 (<http://www.fluxus-engineering.com/>) et en utilisant l'algorithme de jonction-médiane (« median-joining network ») développé par Bandelt (Bandelt *et al.*, 1999). Cet algorithme combine l'algorithme de Kruskal de « minimum spanning tree » (Kruskal, 1956) et l'algorithme heuristique de Farris de maximum de parsimonie (Farris, 1970).

L'analyse phylogénétique a été menée sur chaque portion de gène séparément en utilisant le maximum de vraisemblance (ML). La méthode de ML (Felsenstein, 1993) tente d'inférer un arbre phylogénétique en trouvant celui qui maximise la probabilité d'observer les données X (alignement de séquences) sous les hypothèses évolutives T d'un modèle d'évolution prédéterminé. Ainsi, la vraisemblance de l'arbre $Pr(X/T)$ est égale au produit des

probabilités d'existence de cet arbre calculé pour chaque site i de l'alignement de séquences $Pr(X_i|T)$:

$$Pr(X|T) = \prod_{i=1}^N Pr(X_i|T), \text{ N étant le nombre de sites.}$$

Afin de tester la solidité des nœuds de l'arbre obtenu en ML, on utilise la méthode des bootstraps. Cette méthode crée des pseudo-matrices de données en tirant des sites au hasard avec remise, à partir de l'alignement de séquences. La solidité d'un nœud correspond à la probabilité de retrouver ce même nœud dans les arbres construits à partir des pseudo-matrices. Afin de faire apparaître les longueurs de branches, l'arbre représenté est celui montrant la vraisemblance la plus forte. La méthode de maximum de vraisemblance a été réalisée avec le logiciel MEGA 5 (Tamura *et al.*, 2007) avec 100000 bootstraps.

2.3.6 Mesure du déséquilibre de liaison au sein des gènes

Nous avons calculé le déséquilibre de liaison (DL) global pour toutes les portions de gènes étudiées sur l'échantillon global et les sous-échantillons. Nous avons choisi d'utiliser, comme mesure du DL, le R^2 (Hill & Robertson, 1968) qui mesure la corrélation entre allèles et qui est connu comme étant la mesure du DL la moins sensible à la petite taille des échantillons et à la faible fréquence des allèles (Morton *et al.*, 2001; Weiss & Clark, 2002). L'analyse du DL a été réalisée à l'aide de DNAsP v. 5.10 (Librado & Rozas, 2009), en ne conservant que les sites informatifs et en testant la significativité des valeurs de DL avec un test exact de Fisher adapté aux petits échantillons additionné d'une correction de Bonferroni afin de corriger les valeurs du test et d'éviter les faux positifs. Afin de visualiser ce DL, nous avons utilisé le logiciel TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) (Yu *et al.*, 2006; Bradbury *et al.*, 2007) sur les sites polymorphes et selon les différents échantillonnages.

2.3.7 Étude d'association entre variants alléliques de VTE2-2 et les teneurs en tocophérol au sein du gradient Nord-Sud au Mali

Afin d'effectuer une analyse d'association génotype/phénotype, nous avons au préalable réalisé une étude de la structuration génétique neutre des individus du gradient.

Pour cela, le génotypage de 8 loci microsatellites nucléaires (**Tableau IV.3**) a été effectuée sur l'ensemble des individus de gradient Nord-Sud au Mali. Ces loci ont été développés spécifiquement pour *Vitellaria paradoxa* (Allal *et al.*, 2008) : mCIRVp14, mCIRVp mCIRVp 28, mCIRVp 113, mCIRVp 159, mCIRVp 167, mCIRVp 168, mCIRVp 175 et mCIRVp 181. La méthode utilisée pour l'amplification et le type des marqueurs microsatellites est similaire à celle utilisée par Allal *et al.* (2008).

Tableau IV.3. Nom et caractéristiques des loci microsatellites nucléaires étudiés (Allal *et al.*, 2008). Tm : température d'hybridation ; * : adaptateur M13 (CACGACGTTGTAAAACGAC) permettant le marquage moléculaire.

Nom de locus	Motif répété	Amorces		Tm [C°]	Taille d'allèle [pb]
mCIRVp14	(Ag)4 GG(AG)17	U : *TCTCTTTCATCGTCCTTTG	L : AACACTTGGTCATCATCTTG	58.8	230-274
mCIRVp28	(CA)6	U : *ATTGTTAGTTATGGTTTGG	L : TGATTGCTATTTTGCTTAC	50	162-170
mCIRVp113	(GA)8'(GA)11' (GA)4'(GA)3	U : *ATCGCTAAGAATGGAGGAG	L : TTGAGTTATGAAACGGGAC	50	187-215
mCIRVp159	(TG)5'(GT)6 (GA)10AA(GA)3	U : *CACGAAGAAATATGCTG	L : ATGGATTGCTTTAGGTG	56.1	240-268
mCIRVp167	(GA)17	U : *CTTCAACTTGGTATTAGAGC	L : ATCAGGATAACTTCGCC	50	232-246
mCIRVp168	(CT)12'(CT)5' (CT)2'(CT)4	U : *ATGTTGATTGTGTGTGTG	L : TTGTATGTCTAAGGTTTGG	50	131-169
mCIRVp175	(TG)8	U : *GCATATTGGGAAAAGTAGAC	L : TGAGCATTGAGATAAGAGTG	50	209-219
mCIRVp181	(CA)8	U : *ATATCCTTAATCACTTGCTC	L : GGTGTGAGGTCAGTCTT	50	125-149

Les paramètres classiques de description de la diversité génétique des populations (He, le taux d'hétérozygotie attendu ; Ho, le taux d'hétérozygotie observé ; F_{ST} l'indice de différenciation génétique ; et F_{IS} , l'indice de fixation des populations) ont été calculés sur tous les loci à l'aide des logiciels GENETIX 4.03 (Belkhir *et al.*, 2004) et FSTAT v. 2.9.3.2 (Goudet, 2001). La structuration génétique des populations, sans a priori avec la fonction de vraisemblance, a été estimée par une analyse Bayésienne grâce au logiciel STRUCTURE (Pritchard *et al.*, 2000). Cette approche permet de regrouper des individus en catégorie répondant aux hypothèses d'équilibre d'Hardy-Weinberg et de panmixie. Le modèle sans admixture a été lancé 5 fois consécutives avec 200000 MCMC (les chaînes de Markov avec technique de Monte Carlo pour calculer les probabilités postérieures) et 50000 « burn-in-period ». Le nombre postérieur de clusters a été évalué à l'aide de l'approche développée par Evanno (Evanno *et al.*, 2005).

Pour réaliser l'étude d'association entre le phénotype et le génotype nous avons utilisé les résultats d'analyses chimiques de la teneur en tocophérols alpha, bêta, gamma et



Chapitre IV : Étude du polymorphisme des gènes candidats

delta présentés dans le Chapitre III, et la variabilité génétique observée pour le gène *VTE2-2* au sein des sites du gradient Nord-Sud au Mali. L'objectif de cette démarche est d'évaluer l'importance du gène candidat sélectionné par l'étude des corrélations entre ses variants alléliques et la variation phénotypique. Pour ce faire nous avons utilisé le logiciel TASSEL (Yu *et al.*, 2006; Bradbury *et al.*, 2007). Ce logiciel utilise un modèle linéaire généralisé (procédure GLM) d'analyse permettant d'associer une variable réponse à plusieurs facteurs dont une partie a effet fixe et une partie a effet aléatoire. Ainsi, nous pouvons associer les phénotypes à un polymorphisme de séquence explicatif, en contrôlant l'effet de la structure génétique des populations comme co-variable. Ce modèle permet d'éviter les faux positifs dus à la structure des populations, grâce à l'utilisation d'une Q-matrice basée sur la structure génétique des individus. Nous avons utilisé un premier modèle tenant compte de la structure des populations dont l'équation s'écrit :

$$Y = \mu + L_{VTE2-2} + Q + \varepsilon$$

où Y est la teneur en tocophérol (α , β , γ ou δ), μ est la moyenne des teneurs en tocophérols (α ou γ) ; L_{VTE2-2} , l'effet du variant allélique ; Q, la matrice de structure ; et ε , l'effet résiduel du modèle.

Un deuxième modèle ne tenant pas compte de la structure a également été utilisé :

$$Y = \mu + L_{VTE2-2} + \varepsilon$$

où Y est la teneur en tocophérol (α , β , γ ou δ), μ est la moyenne des teneurs en tocophérols ; L_{VTE2-2} , l'effet du variant allélique ; et ε , l'effet résiduel du modèle.

3 RÉSULTATS

3.1 Étude cytogénétique et comptage chromosomique

L'étude cytogénétique de plusieurs accessions de *Vitellaria paradoxa* ssp. *paradoxa* et ssp. *nilotica* révèle un niveau de ploïdie identique pour les deux sous-espèces (2n) avec un total de 26 chromosomes, soit 13 paires (**Figure IV.7**). L'observation du caryotype révèle également que la taille des paires de chromosomes est quasi identique, sauf pour deux paires qui semblent plus longues que les autres.

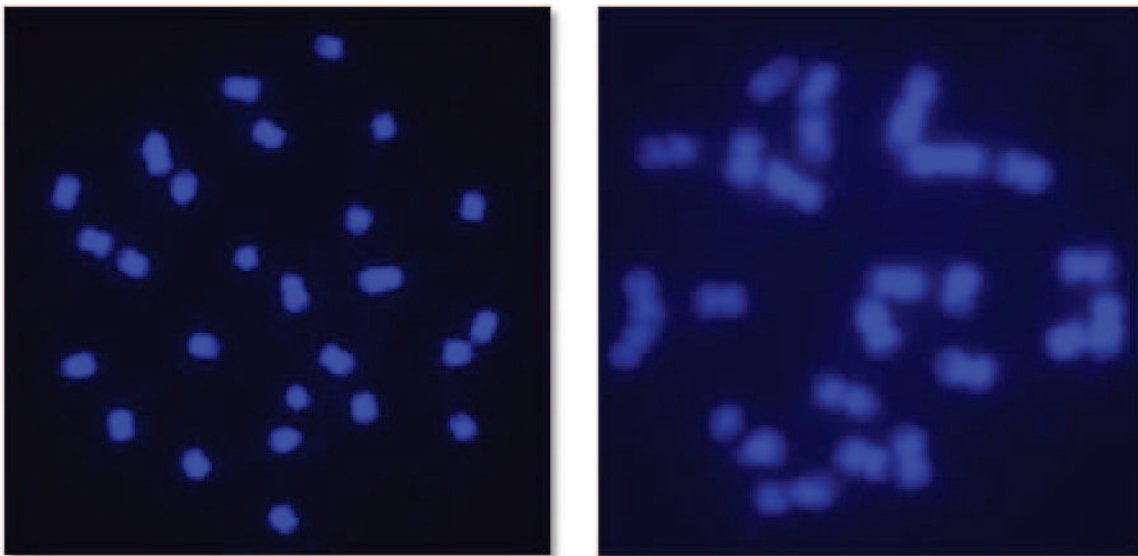


Figure IV.7. Observation de chromosomes en métaphase. A : accession de *Vitellaria paradoxa* ssp. *nilotica* ; B : accession de *Vitellaria paradoxa* ssp. *paradoxa*.

3.2 Étude du gène SAD chez *Vitellaria paradoxa*

3.2.1 Structure et séquence du gène SAD

Le séquençage de l'ADN complémentaire du gène *SAD* révèle pour la sous-espèce *paradoxa* deux versions d'ARNm de respectivement 1185 et 1165 pb. La différence entre ces deux ARNm consiste essentiellement en une délétion d'une partie de l'exon 1 (20 pb). Ce type d'observation est parfois dû à des versions alternatives d'épissage de l'ARN pré-messenger (Black, 2003; Matlin *et al.*, 2005). Cependant, le séquençage de l'ADN génomique révèle qu'il s'agit de deux copies extrêmement proches (*SAD1* et *SAD2*) mais distinctes du gène *SAD* chez *Vitellaria paradoxa* ssp. *paradoxa*, dont les séquences nucléotidiques (*SAD1* : 2605 pb et *SAD2* : 2585 pb) sont homologues à 99% (**Figure IV.8**).

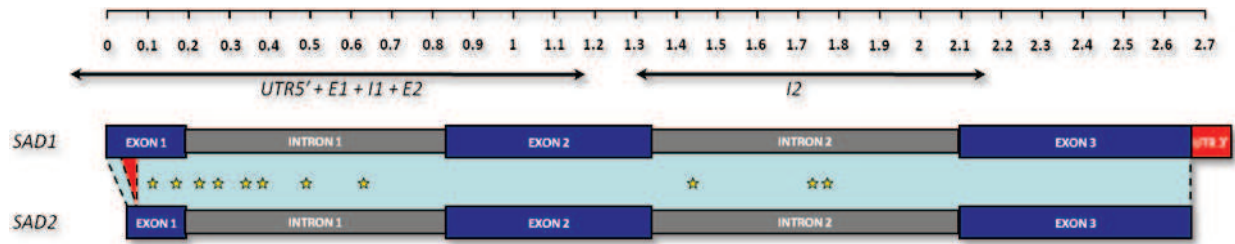


Figure IV.8. Structure et colinéarité des copies *SAD1* et *SAD2* au sein de la sous-espèce *paradoxa*. Le triangle rouge représente la délétion observée pour *SAD2* et les étoiles jaunes les mutations copies spécifiques. Les flèches noires délimitent les parties des deux copies utilisées dans les analyses suivantes. L'échelle graduée est exprimée en kb.

La traduction *in silico* de la copie *SAD1* révèle une protéine de 395 acides aminés (Figure IV.9) homologues à 89% à la séquence protéique de *SAD* obtenue chez *Jatropha curcas* (Tong *et al.*, 2006). De la traduction *in silico* de l'ARNm de la copie *SAD2* chez la sous-espèce *paradoxa* résulte une protéine tronquée (Figure IV.9). Ainsi, la délétion observée dans la copie *SAD2* entraîne un décalage dans la phase de lecture révélant de multiples codons « stops ». En revanche, au sein de la sous-espèce *nilotica*, le séquençage de la copie de gène *SAD2* révèle l'absence de la délétion. En conséquence, la traduction *in silico* de *SAD2* au sein des individus *nilotica*, entraîne l'obtention d'une protéine normale (non tronquée). En conclusion, il semble que les deux copies du gène *SAD* puissent entraîner la traduction d'une protéine fonctionnelle chez les individus de la sous-espèce *nilotica*, tandis qu'au sein de la sous-espèce *paradoxa*, *SAD2* serait non fonctionnelle (Annexe IV.2).

Copie de gène	Position	Séquence protéique
SAD1	001	■ RGMKNQLCS RRISCAPLFD LFILRRRKEK IKFSFLQYTO REVDNLKKPF TPPREVHLQV THSMPPQKIE IFKNLEDWAE QNVLVHLKPV EKCWQPDQFL
	101	PDPASDGFHD QVNELRKRAN EIPDDYFIVL VGDMITEEAL PTYQTMNLNLT DGVRDEXGAS LTSWARWTRA WTAENNRHGD LLNKLYLMSG RVDMKQIEKT
	201	IQYLIQSGMD PRTEENPYLG FIYTSFQERA TFISHGNTAR LAKEHGDIKL AQICGTIASD EKRKETAYTK IVEKLFEIDP DGTVVALADM MKKKISMPAH
	301	LMYDGRDDNL FDHFSAVAQR LGVYTAKDYA DILEFLVGRW RVEEVTGLSA EGRKAQDYVC GLPPRIRRL ERSQSKAKQT RIIPFSWIFD REVQL ■
SAD2	001	■ RGMKNQLCS NFLTFSSSEE ERK ■
	101
	201
	301

Figure IV.9. Protéines *SAD1* et *SAD2* chez un individu de la sous espèce *paradoxa*, déduites par traduction *in silico* des séquences d'acide nucléique. En vert est représenté le codon « start », en rouge le codon « stop » et en pointillés la partie tronquée de la protéine *SAD2*

Chapitre IV : Étude du polymorphisme des gènes candidats

Tableau IV.4. Diversité et sélection des gènes *SAD1* et *SAD2* considérant différents échantillons : Total = toutes les séquences disponibles de l'aire naturelle ; Ouest = séquences des individus d'Afrique de l'Ouest ; Centre = séquences d'Afrique Centrale ; Est = séquences d'Afrique de l'Est. N seq : nombre de séquences ; N site : nombre de nucléotides ; S : nombre de sites en ségrégation ; Hap : nombre d'haplotypes ; Hd : diversité haplotypique (Nei, 1987) ; π : estimateur de la diversité nucléotidique de Nei (Nei, 1987) ; θ_w : estimateur de la diversité nucléotidique de Watterson (Watterson, 1975) ; D : valeur du D de Tajima (Tajima, 1989) ; D* et F* : valeur des D* et F* de Fu et Li (Fu & Li, 1993). # : P<0.05 ; * : P< 0.01 ; ** : P< 0.001.

Gènes		<i>SAD1</i>					<i>SAD2</i>				
Région		<i>SAD1_5'</i>				<i>SAD1_3'</i>	<i>SAD2_5'</i>				<i>SAD2_3'</i>
Populations	Domaine Position	UTR5' [-42↔0]	Exons 1 et 2 [1↔113] [757↔1077]	Intron 1 [114↔756]	Total [-42↔1077]	Intron 2 [1238↔2057]	UTR5' [-42↔0]	Exons 1 et 2 [1↔113] [757↔1077]	Intron 1 [114↔756]	Total [-42↔1077]	Intron 2 [1238↔2057]
Total	N seq	32	32	32	32	52	31	31	31	31	52
	N sites	42	433	644	1119	820	42	433	643	1118	821
	S	0	4	24	28	19	0	15	33	48	49
	Hap	1	5	19	19	19	1	13	21	23	39
	Hd	0	0.341	0.931	0.931	0.851	0	0.822	0.963	0.978	0.98
	π	0	0.00085	0.00428	0.00279	0.00183	0	0.00424	0.0053	0.0047	0.00749
	θ_w	-	0.00229	0.00928	0.00622	0.00515	-	0.00909	0.01285	0.01094	0.01327
	D	-	-1.5916 #	-1.8901 *	-1.962 *	-2.027 *	-	-1.786 #	-2.1256 *	-2.1117 *	-1.4933
	D*	-	-0.999	-3.2497 *	-3.085 *	-1.836	-	-2.8355 *	-2.3613 #	-2.7388 *	-3.3999 **
	F*	-	-1.3628	-3.3093 **	-3.2057 *	-2.2564 #	-	-2.9399 *	-2.6946 *	-2.9907 *	-3.2226 *
Ouest	N seq	24	24	24	24	28	23	23	23	23	28
	N sites	42	433	644	1119	820	42	433	643	1118	821
	S	0	4	21	25	12	0	12	28	40	27
	Hap	1	5	17	17	12	1	11	17	19	19
	Hd	0	0.438	0.953	0.953	0.831	0	0.735	0.953	0.98	0.972
	π	0	0.00112	0.00376	0.0026	0.00168	0	0.00272	0.00534	0.00415	0.00653
	θ_w	-	0.00247	0.00876	0.00599	0.00377	-	0.00787	0.0118	0.00987	0.00857
	D	-	-1.4953	-2.089*	-2.107*	-1.8402 *	-	-2.2693 **	-2.0733*	-2.244 **	-0.8693
	D*	-	-0.856	-3.0614**	-2.8411*	-1.2863	-	-3.2727 **	-2.3271#	-2.8446 *	-1.5604
	F*	-	-1.1998	-3.2319**	-3.0605*	-1.7036	-	-3.4635 **	-2.6303 *	-3.1133 *	-1.5762
Centre	N seq	-	-	-	-	16	-	-	-	-	16
	N sites	-	-	-	-	820	-	-	-	-	821
	S	-	-	-	-	6	-	-	-	-	22
	Hap	-	-	-	-	6	-	-	-	-	12
	Hd	-	-	-	-	0.742	-	-	-	-	0.967
	π	-	-	-	-	0.00145	-	-	-	-	0.00662
	θ_w	-	-	-	-	0.00221	-	-	-	-	0.00811
	D	-	-	-	-	-1.1829	-	-	-	-	-0.7417
	D*	-	-	-	-	-0.047	-	-	-	-	-0.8768
	F*	-	-	-	-	-0.4069	-	-	-	-	-0.9675
Est	N seq	8	8	8	8	8	8	8	8	8	8
	N sites	42	433	644	1119	820	42	433	643	1118	821
	S	0	0	1	1	2	0	1	4	5	8
	Hap	1	1	2	2	3	1	2	4	4	7
	Hd	0	0	0.25	0.25	0.607	0	0.429	0.821	0.821	0.964
	π	0	0	0.00039	0.00022	0.00083	0	0.00099	0.00239	0.00176	0.00266
	θ_w	-	-	0.0006	0.00034	0.00094	-	0.00089	0.0024	0.00172	0.00377
	D	-	-	-1.0548	-1.0548	-0.4479	-	0.3335	-0.0198	0.0844	-1.4213
	D*	-	-	-1.1264	-1.1264	-0.1493	-	0.8878	0.5681	0.7471	-1.4435
	F*	-	-	-1.2035	-1.2035	-0.2379	-	0.8253	0.475	0.6539	-1.5911



3.2.2 Étude du polymorphisme des gènes *SAD1* et *SAD2* chez

Vitellaria paradoxa

Le **Tableau IV.4** présente l'ensemble des résultats obtenus pour l'analyse du polymorphisme des séquences *SAD1* et *SAD2*. Ne disposant pas d'un échantillonnage similaire pour les parties 5' (UTR5' + Exon 1 + Intron 2 + Exon 2) et 3' (Intron 2) des deux copies de gène *SAD1* et *SAD2*, nous avons étudié ces fragments séparément. Nous parlerons donc de *SAD1_5'* et *SAD2_5'* pour désigner la partie 5' (UTR5' + Exon 1 + Intron 2 + Exon 2), et de *SAD1_3'* et *SAD2_3'* pour désigner la partie 3' étudiée (Intron 2).

Les résultats montrent que la copie *SAD1* est moins polymorphe que la séquence *SAD2*. Ainsi, dans la partie 5' (de -42 pb à 1077 pb) pour l'échantillon total, on observe pour *SAD1_5'* 28 sites en ségrégations résultant en 19 haplotypes avec une diversité moyenne de $Hd_{SAD1_5'} = 0.931$. Pour la même région mais au sein de la copie *SAD2_5'*, 48 sites polymorphes sont détectés avec 23 haplotypes, la diversité haplotypique $Hd_{SAD2_5'} = 0.978$ (**Tableau IV.4**). Les résultats révèlent que la valeur des estimateurs π et θ_w est deux fois plus forte au sein de *SAD2_5'* ($\pi_{SAD2_5'} = 0.0047$; $\theta_w_{SAD2_5'} = 0.01094$) que pour la copie *SAD1* ($\pi_{SAD1_5'} = 0.00279$; $\theta_w_{SAD1_5'} = 0.00622$) (**Tableau IV.4**).

De la même façon, on observe une diversité accrue au sein de *SAD2_3'* par rapport à *SAD1_3'*. Ainsi, on détecte au sein de *SAD1_3'* seulement 19 sites polymorphes, avec 19 haplotypes, $Hd_{SAD1_3'} = 0.851$, $\pi_{SAD1_3'} = 0.00183$ et $\theta_w_{SAD1_3'} = 0.00515$; contre 49 SNPs au sein de *SAD2_3'*, avec 36 haplotypes, $Hd_{SAD2_3'} = 0.98$, $\pi_{SAD2_3'} = 0.00749$ et $\theta_w_{SAD2_3'} = 0.01327$ (**Tableau IV.4**).

On observe que les exons de la copie *SAD1* présentent une diversité haplotypique très faible ($Hd_{SAD1_Exons} = 0.341$) par rapport à la diversité globale de *SAD1_5'* (**Tableau IV.4**). À l'inverse, les exons de la copie *SAD2* présentent une diversité haplotypique ($Hd_{SAD2_Exons} = 0.822$) quasiment similaire à la diversité globale de *SAD2_5'* (**Tableau IV.4**).

Au sein des sous-échantillons de populations étudiés, on observe que les patrons de variations à l'Ouest sont semblables à ceux observés sur l'échantillon total pour les deux copies *SAD1* et *SAD2* (**Tableau IV.4**). C'est également le cas pour le sous-échantillon



d'Afrique Centrale, pour lequel l'étude de l'Intron 2 des deux copies du gène *SAD* révèle une diversité plus forte pour *SAD2* (**Tableau IV.4**). Cependant pour le sous-échantillon d'Afrique de l'Est, on observe que la diversité est globalement plus faible, et ce pour les deux copies *SAD1* et *SAD2*. De plus, contrairement au reste de l'échantillon, on observe au sein de *SAD2* une diversité très faible des exons pour l'échantillon d'Afrique de l'Est (**Tableau IV.4**).

3.2.3 Étude du déséquilibre de liaison (DL) au sein des gènes *SAD1* et *SAD2*

L'analyse du DL au sein des gènes *SAD1* et *SAD2* montre des valeurs globales de R^2 très faibles et non significatives (**Figure IV.10** et **IV.11**). Cependant le nombre de paires de nucléotides testées est très faible du fait du faible polymorphisme détecté au sein des fragments de gène étudiés. On note cependant quelques valeurs significatives pour certaines paires de nucléotides testées. Ainsi les valeurs représentées en rouge sont non seulement significatives au regard du test d'exactitude de Fisher, mais également suite à la correction de Bonferroni (**Figure IV.10** et **IV.11**).

L'observation de la variation du déséquilibre de liaison au sein du gène *SAD1* (**Figure IV.10**) montre des valeurs de DL fortes pour des paires particulièrement éloignées de nucléotides, au sein de l'échantillon global. De rares paires de nucléotides montrent des DL significatifs au sein du sous-échantillon de l'Ouest, très peu pour le sous-échantillon d'Afrique Centrale, et aucun pour l'échantillon d'Afrique de l'Est (**Figure IV.10**).

On observe pour le gène *SAD2* (**Figure IV.11**) un déséquilibre de liaison très faible, mais supérieur à celui mis en évidence pour *SAD1* (**Figure IV.10**). En particulier, on note des valeurs de R^2 de DL très forts pour un nombre plus important de paires de nucléotides répartis de façon relativement homogène au sein du gène. On observe un DL de liaison plus faible au sein des sous-échantillons d'Afrique de l'Ouest et d'Afrique Centrale que pour celui d'Afrique de l'Est (**Figure IV.11**), seules très peu de valeurs sont significatives au regard du test d'exactitude de Fisher.

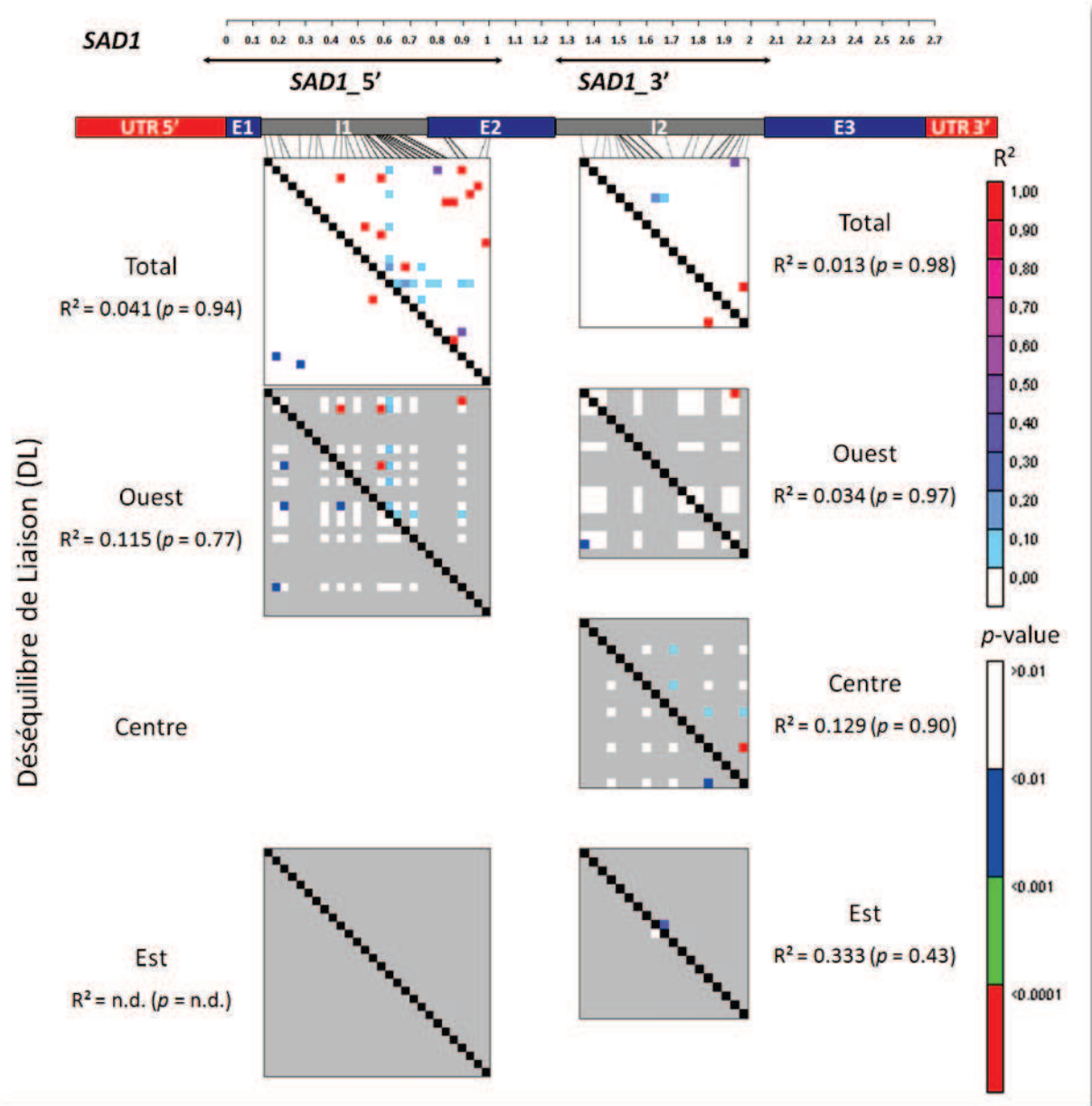


Figure IV.10. Déséquilibre de liaison au sein du gène *SAD1* considérant différents échantillons : Total = toutes les séquences disponibles de l'aire naturelle ; Ouest = séquences des individus d'Afrique de l'Ouest ; Centre = séquences d'Afrique Centrale ; Est = séquences d'Afrique de l'Est. Les matrices triangulaires supérieures indiquent les valeurs des R^2 entre paires de nucléotides. Les matrices triangulaires inférieures indiquent les degrés de significativité des valeurs de DL obtenu avec le test d'exactitude de Fisher corrigé par Bonferroni. L'échelle graduée au dessus du gène est exprimée en kb.

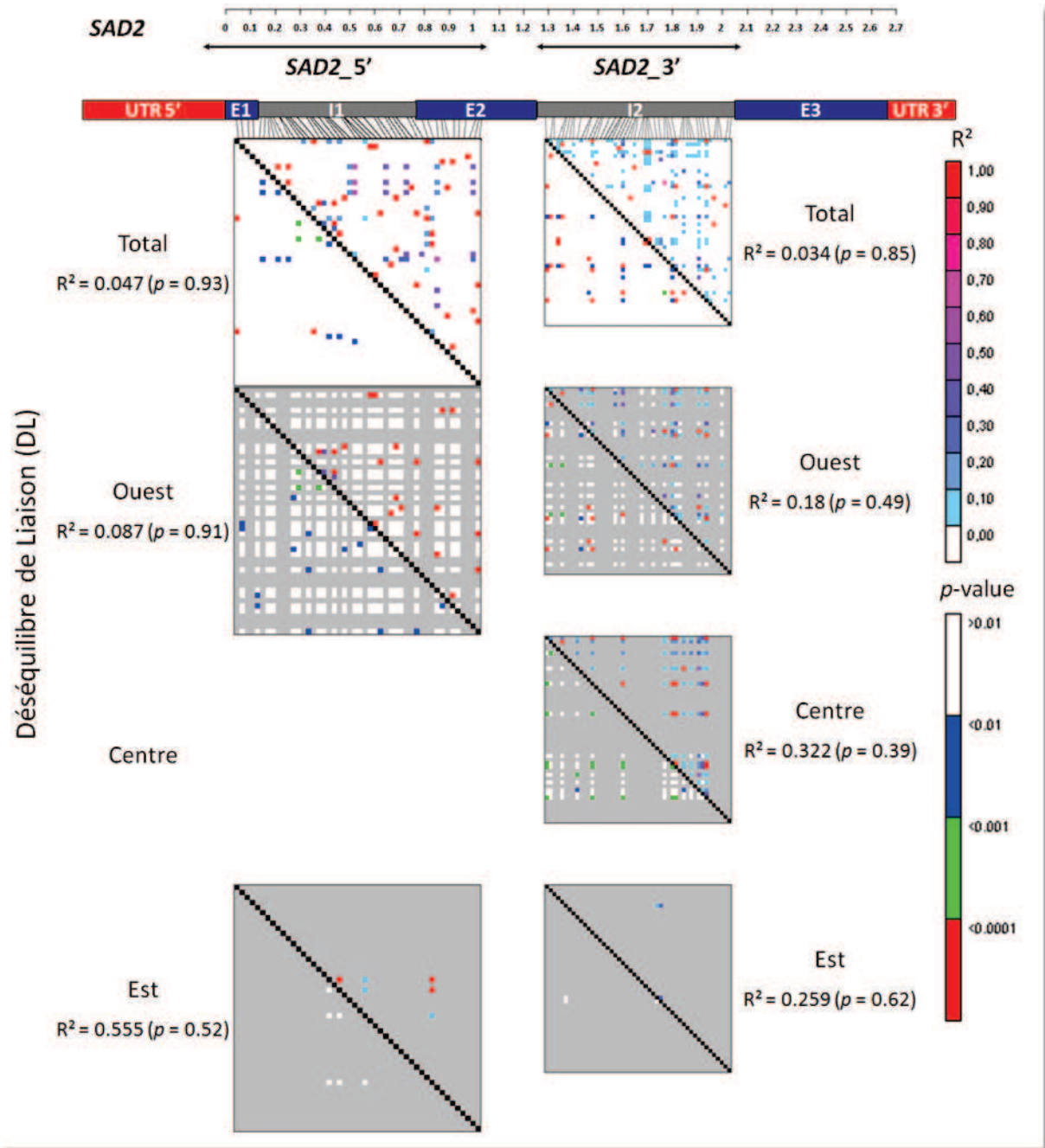


Figure IV.11. Déséquilibre de liaison au sein du gène *SAD2* considérant différents échantillons : Total = toutes les séquences disponibles de l'aire naturelle ; Ouest = séquences des individus d'Afrique de l'Ouest ; Centre = séquences d'Afrique Centrale ; Est = séquences d'Afrique de l'Est). Les matrices triangulaires supérieures indiquent les valeurs des R^2 entre paires de nucléotides. Les matrices triangulaires inférieures indiquent les degrés de significativité des valeurs de DL obtenu avec le test d'exactitude de Fisher corrigé par Bonferroni. L'échelle graduée au dessus du gène est exprimée en kb.

3.2.4 Sélection des gènes *SAD1* et *SAD2* chez *Vitellaria paradoxa*

Sur l'échantillon total, les tests de neutralité suggèrent globalement une sélection positive pour les deux copies *SAD1* et *SAD2* (**Tableau IV.4**). En effet, *SAD1* et *SAD2* présentent des excès significatifs de sites polymorphes rares, révélés par des D de Tajima négatifs au seuil 2%, et des excès significatifs de sites polymorphes récents, révélés par des D^* et des F^* de Fu et Li négatifs au seuil 2% (**Tableau IV.4**). Ces résultats sont cependant à prendre avec prudence, car ces statistiques sont fortement impactées par les phénomènes démographiques. Un tel excès de singletons/mutations rares peut être ainsi généré par des phénomènes d'expansion démographique. Ce point est discuté plus loin dans la section Discussion de ce Chapitre.

Dans le sous-échantillon d'Afrique de l'Ouest, les tests de neutralité sont également négatifs et significatifs pour les gènes *SAD1* et *SAD2*, indiquant également un excès de variants alléliques rares au sein de la région Ouest (**Tableau IV.4**). Pour cet échantillon on observe que les tests sont cependant non significatifs au sein de l'Exon 2 du gène *SAD1*. En revanche, on observe des valeurs très significatives pour D , D^* et F^* au sein de cette même région génique pour la copie non-fonctionnelle *SAD2* (**Tableau IV.4**).

Dans les sous-échantillons d'Afrique de l'Est et d'Afrique Centrale, aucun des tests de neutralité D , D^* et F^* ne sont significatifs, que se soit au sein des régions codantes ou non-codantes. Ce résultat suggère donc que les patrons de variation des indices de neutralité sélective ne sont pas homogènes entre les différents échantillons (**Tableau IV.4**). Cependant, ce résultat peut également résulter de la densité de l'échantillonnage plus faible au sein de l'Afrique de l'Est et de l'Afrique Centrale (respectivement 8 et 16 séquences ; **Tableau IV.1**) que pour l'échantillon d'Afrique de l'Ouest (23 à 28 séquences).

L'évolution des indices D , D^* et F^* pour les deux copies *SAD1* (**Figure IV.12**) et *SAD2* (**Figure IV.13**) montre que les tests de neutralité ne sont pas homogènes le long des portions des deux gènes étudiés, avec plus de valeurs significatives au sein des introns. Ainsi, confirmant les résultats apportés par le **Tableau IV.4**, on note que seul le sous-échantillon Ouest montre des valeurs significatives pour les tests de neutralité. La « sliding-window » révèle en outre pour le gène *SAD1* beaucoup de valeurs significatives au sein des introns

(Figure IV.12), tandis que les valeurs significatives sont mieux réparties sur le long des fragments pour le gène *SAD2* (Figure IV.13).

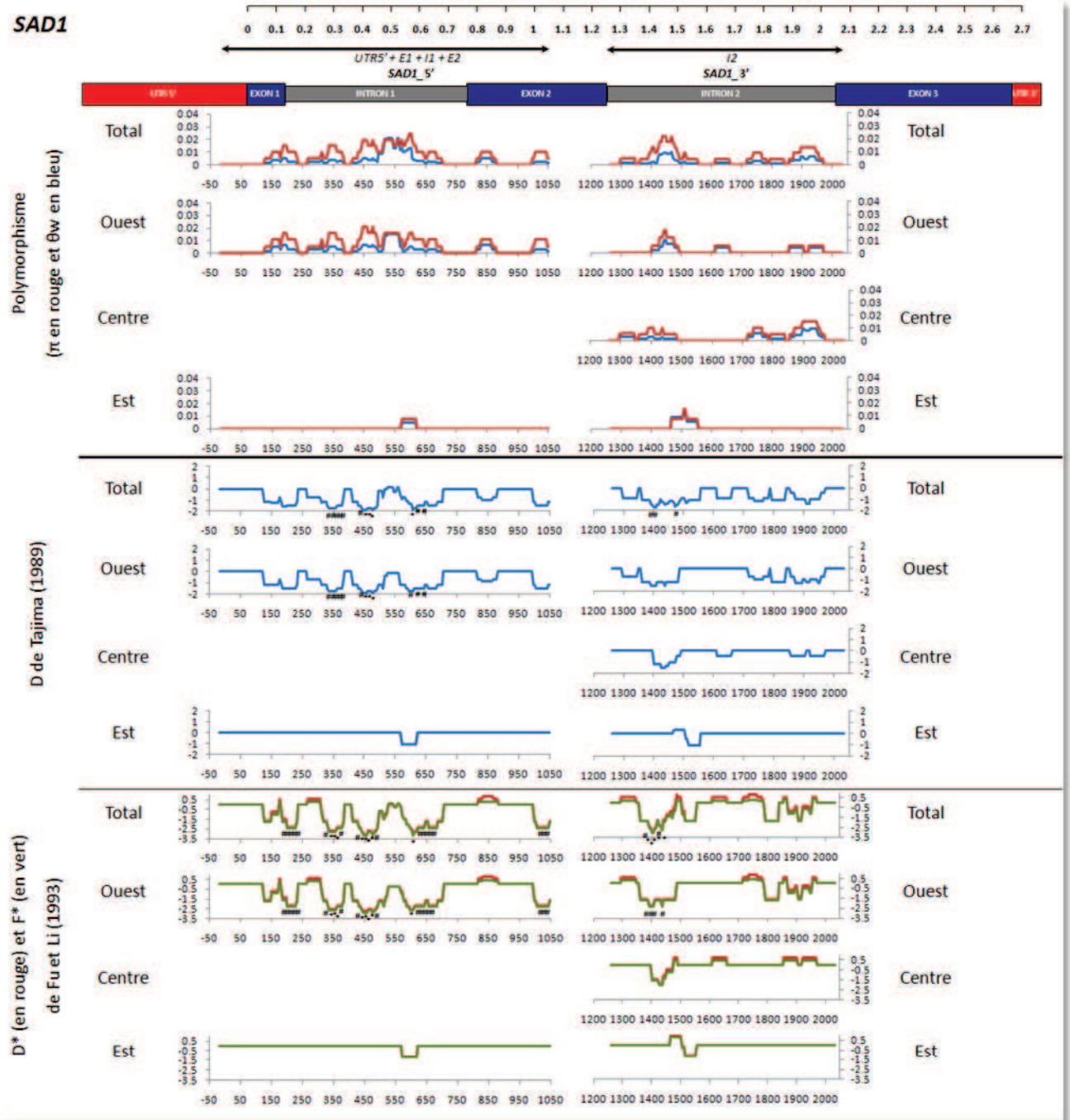


Figure IV.12. Évolution de la diversité et des indices de sélection du gène *SAD1* en utilisant la méthode de la « sliding window » du logiciel DNAsP v. 5.10 (Librado & Rozas, 2009) considérant différents échantillons : Total = toutes les séquences disponibles de l'aire naturelle ; Ouest = séquences des individus d'Afrique de l'Ouest ; Centre = séquences d'Afrique Centrale ; Est = séquences d'Afrique de l'Est. π : estimateur de la diversité nucléotidique de Nei (Nei, 1987); θ_w : estimateur de la diversité nucléotidique de Watterson (Watterson, 1975); D : valeur du D de Tajima (Tajima, 1989) ; D* et F* : valeur des D* et F* de Fu et Li (Fu & Li, 1993). # : $P < 0.05$; * : $P < 0.01$. L'échelle graduée au dessus du gène est exprimée en kb.

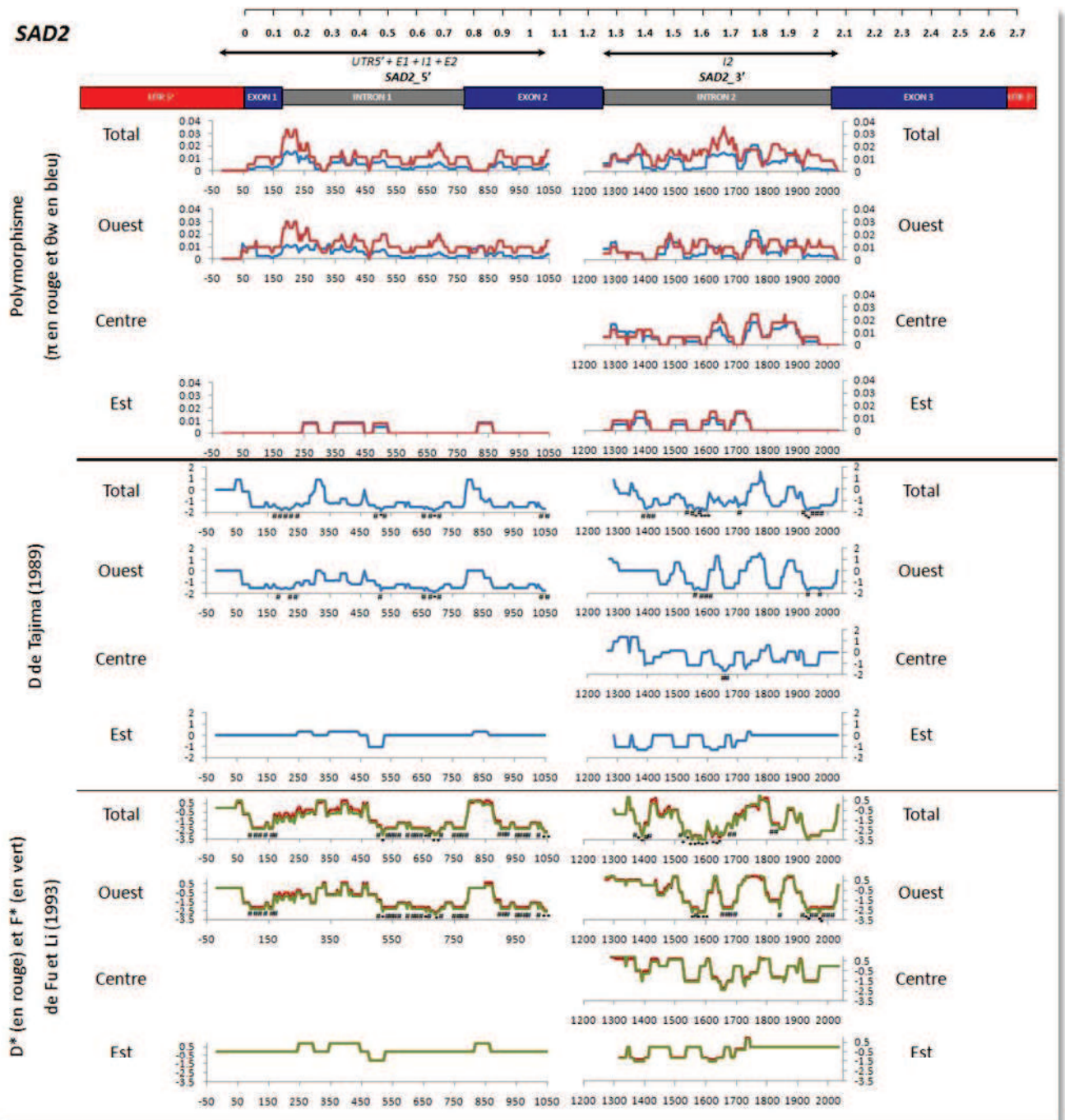


Figure IV.13. Évolution de la diversité et des indices de sélection du gène *SAD2* en utilisant la méthode de la « sliding window » du logiciel DNAsP v. 5.10 (Librado & Rozas, 2009) considérant différents échantillons : Total = toutes les séquences disponibles de l'aire naturelle ; Ouest = séquences des individus d'Afrique de l'Ouest ; Centre = séquences d'Afrique Centrale ; Est = séquences d'Afrique de l'Est. π : estimateur de la diversité nucléotidique de Nei (Nei, 1987); θ_w : estimateur de la diversité nucléotidique de Watterson (Watterson, 1975); D : valeur du D de Tajima (Tajima, 1989) ; D* et F* : valeur des D* et F* de Fu et Li (Fu & Li, 1993). # : $P < 0.05$; * : $P < 0.01$. L'échelle graduée au dessus du gène est exprimée en kb.

3.2.5 Différentiation génétique et réseaux des haplotypes des gènes *SAD1* et *SAD2*

Afin de quantifier le niveau de différenciation génétique entre les populations au sein des loci *SAD1* et *SAD2*, nous avons calculé les indices de F_{ST} (Hudson *et al.*, 1992) et de G_{ST} (Nei, 1973), de chaque copie *SAD1* et *SAD2*, pour chaque fragment de gène, à l'aide de DnaSP v. 5.10 (Rozas and Rozas, 1999).

Tableau IV.5. Niveaux de différenciation génétique F_{ST} (en noir) et G_{ST} (en rouge) entre les populations Ouest (Afrique de l'Ouest), Centre (Afrique Centrale) et Est (Afrique de l'Est) pour les différents fragments des gènes *SAD1* et *SAD2*. Les valeurs sont significatives au seuil 5%. n.m. : valeur non mesurée.

<i>SAD1_5'</i>	Population	Ouest	Centre	Est
	Ouest	-	n.m.	0.062
	Centre	n.m.	-	n.m.
	Est	0.029	n.m.	-
<i>SAD1_3'</i>	Population	Ouest	Centre	Est
	Ouest	-	0.033	0.076
	Centre	0.008	-	0.111
	Est	0.048	0.026	-
<i>SAD2_5'</i>	Population	Ouest	Centre	Est
	Ouest	-	n.m.	0.134
	Centre	n.m.	-	n.m.
	Est	0.144	n.m.	-
<i>SAD2_3'</i>	Population	Ouest	Centre	Est
	Ouest	-	0.121	0.210
	Centre	0.010	-	0.325
	Est	0.022	0.021	-

Ainsi, le **Tableau IV.5** présente les indices de différenciation génétique entre les populations d'Afrique de l'Ouest, d'Afrique Centrale et d'Afrique de l'Est pour les différents fragments analysés des gènes *SAD1* et *SAD2*. Globalement, on observe des valeurs de F_{ST} et de G_{ST} faibles pour la copie *SAD1* avec des valeurs de $F_{ST} < 0.12$ et de $G_{ST} < 0.05$ (**Tableau IV.5**). Cependant, ces valeurs sont similaires à celles observées avec les microsatellites nucléaires entre les mêmes populations de *Vitellaria paradoxa* (Ouest/Est : $F_{ST} = 0.07$; Ouest/Centre : $F_{ST} = 0.07$; Centre/Est : $F_{ST} = 0.08$) (cf Chapitre II). En revanche, on note une différenciation génétique plus forte sur la base des fragments de *SAD2*, avec des valeurs de

$F_{ST} > 0.12$ (**Tableau IV.5**). Cependant, on observe que les valeurs de G_{ST} restent très faibles avec $G_{ST} < 0.03$ (**Tableau IV.5**).

Les réseaux haplotypiques des fragments *SAD1_5'*, *SAD2_5'*, *SAD1_3'* et *SAD2_3'* sont présentés aux **Figures IV.14**, **IV.15**, **IV.16** et **IV.17**. Pour interpréter les résultats obtenus, nous faisons l'hypothèse que l'haplotype central, majoritaire est l'haplotype ancestral et que les haplotypes en faible fréquence rayonnant à partir de l'haplotype principal sont des haplotypes secondaires.

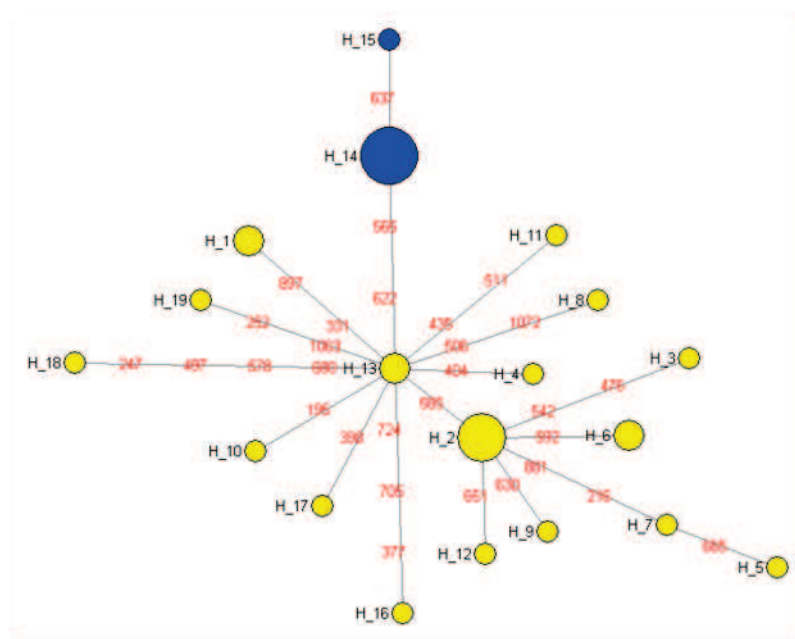


Figure IV.14. Réseau haplotypique de *SAD1_5'* obtenu par la méthode de "median-joining" (Bandelt *et al.*, 1999) grâce au logiciel NETWORK 4.516. La position des mutations est indiquée en rouge sur les branches. Chaque haplotype est représenté par un cercle de diamètre proportionnel à sa fréquence. Les haplotypes d'Afrique de l'Ouest sont en jaune et les haplotypes d'Afrique de l'Est sont en bleu.

Le réseau haplotypique obtenu pour *SAD1_5'* ne permet cependant pas de mettre en évidence un haplotype très majoritaire (**Figure IV.14**). On observe en outre que la divergence entre les 19 haplotypes mis en évidence, est faible avec au maximum 8 nucléotides divergents entre les haplotypes H_5 et H_18. On observe également qu'il n'y a pas de partage d'haplotypes entre les groupes Est et Ouest, et que les deux haplotypes de l'Est sont phylogénétiquement très proches avec seulement une différence sur un seul nucléotide. Enfin, on remarque que les haplotypes H_2 et H_13 divergents d'une seule paire de base, sont centraux et à l'origine de la quasi-totalité des autres haplotypes.

Le réseau haplotypique obtenu pour *SAD2_5'* (**Figure IV.15**) est très différent de celui de la copie *SAD1* (**Figure IV.14**). En effet, on constate d'une part une grande divergence des

haplotypes d'Afrique de l'Est par rapport à ceux d'Afrique de l'Ouest, avec une divergence de 23 paires de bases au minimum (**Figure IV.15**). De plus, on observe que les longueurs de branches (le nombre de mutation) entre les haplotypes de l'Ouest sont relativement importantes, contrairement à la faible divergence observée entre les haplotypes de l'Est. L'haplotype H_11 semble central au sein du réseau d'haplotypes, et semble être à l'origine de nombreux allèles (**Figure IV.15**).

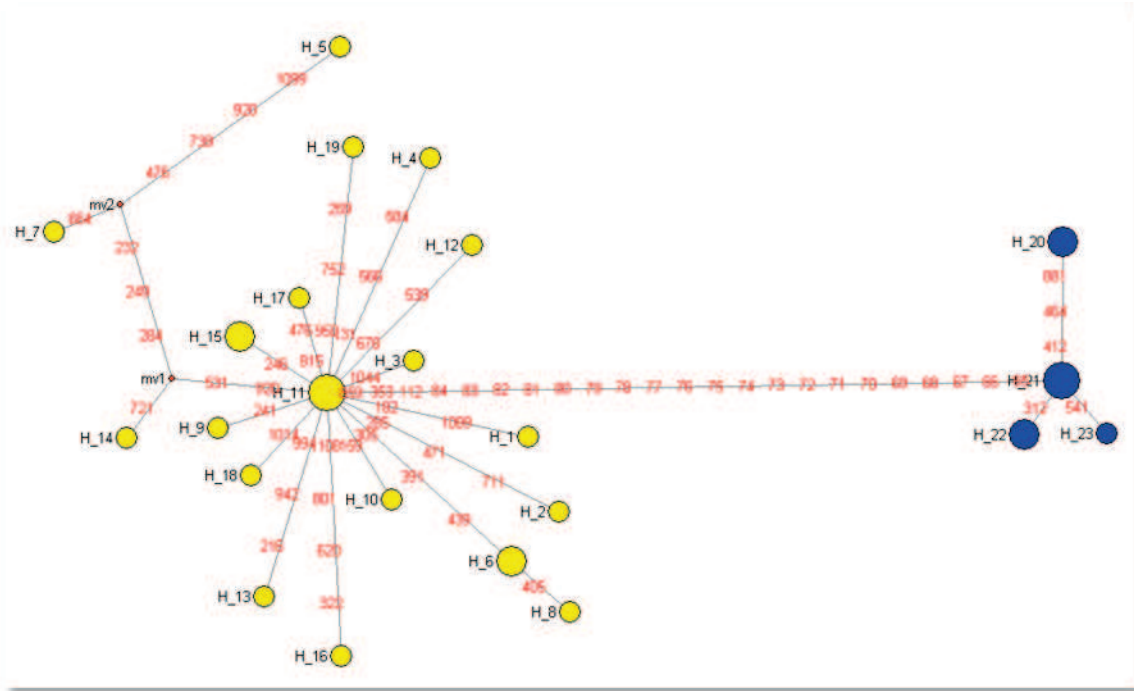
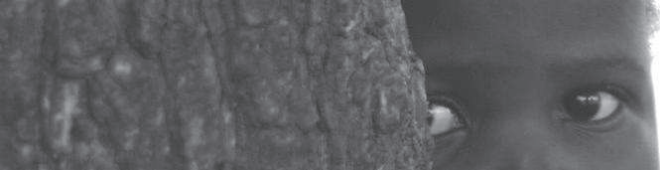


Figure IV.15. Réseau haplotypique de *SAD2_5'* obtenu par la méthode de "median-joining" (Bandelt *et al.*, 1999) grâce au logiciel NETWORK 4.5.16. Les points rouges représentent des haplotypes de jonction manquants. La position des mutations est indiquée en rouge sur les branches. Chaque haplotype est représenté par un cercle de diamètre proportionnel à sa fréquence. Les haplotypes d'Afrique de l'Ouest sont en jaune et les haplotypes d'Afrique de l'Est sont en bleu.

La topologie proposée par le réseau haplotypique du fragment *SAD1_3'* (**Figure IV.16**) ressemble à celle observée pour le réseau haplotypique *SAD1_5'* (**Figure IV.14**). Ainsi, on note le même type de structure en « étoile », caractéristique d'un phénomène de sélection positive ou d'expansion démographique. On observe ici un haplotype central majoritaire (H_6) partagé à la fois par la population d'Afrique de l'Ouest mais également pas la population d'Afrique centrale. On note que trois haplotypes sont partagés entre les deux populations Ouest et Centre (H_3, H_6 et H_9). Cependant, on remarque que les longueurs de branches sont très courtes avec seulement 1 paire de bases de divergence observée entre chacun des haplotypes. On constate enfin, comme pour les marqueurs précédents, que la



Chapitre IV : Étude du polymorphisme des gènes candidats

population Est ne partage aucun de ses haplotypes, qui sont par ailleurs très proches phylogénétiquement les uns des autres.

Le réseau haplotypique obtenu pour *SAD2_3'* (**Figure IV.17**) montre une structure complexe et très éclatée, avec un total de 39 haplotypes, dont un seul est partagé entre les populations d'Afrique centrale et d'Afrique de l'Ouest (H_5). On note que les haplotypes d'Afrique de l'Ouest et d'Afrique Centrale sont très mélangés, et les longueurs de branches entre chaque haplotype sont globalement plus longues que celles observées dans le réseau haplotypique *SAD1_3'* (**Figure IV.16**). En revanche, on note une fois encore que les haplotypes spécifiques d'Afrique de l'Est sont bien regroupés et très proches phylogénétiquement les uns des autres (**Figure IV.17**). Enfin, on remarque que trois haplotypes de jonction sont manquants dans l'échantillonnage.

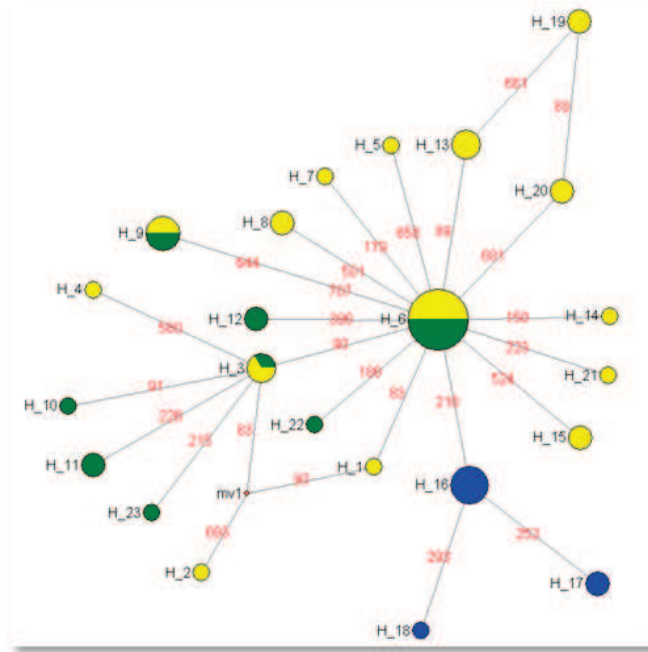


Figure IV.16. Réseau haplotypique de *SAD1_3'* obtenu par la méthode de "median-joining" (Bandelt *et al.*, 1999) grâce au logiciel NETWORK 4.516. Les points rouges représentent des haplotypes de jonction manquants. La position des mutations est indiquée en rouge sur les branches. Chaque haplotype est représenté par un cercle de diamètre proportionnel à sa fréquence. Les haplotypes d'Afrique de l'Ouest sont en jaune, les haplotypes d'Afrique Centrale sont en vert et les haplotypes d'Afrique de l'Est sont en bleu.

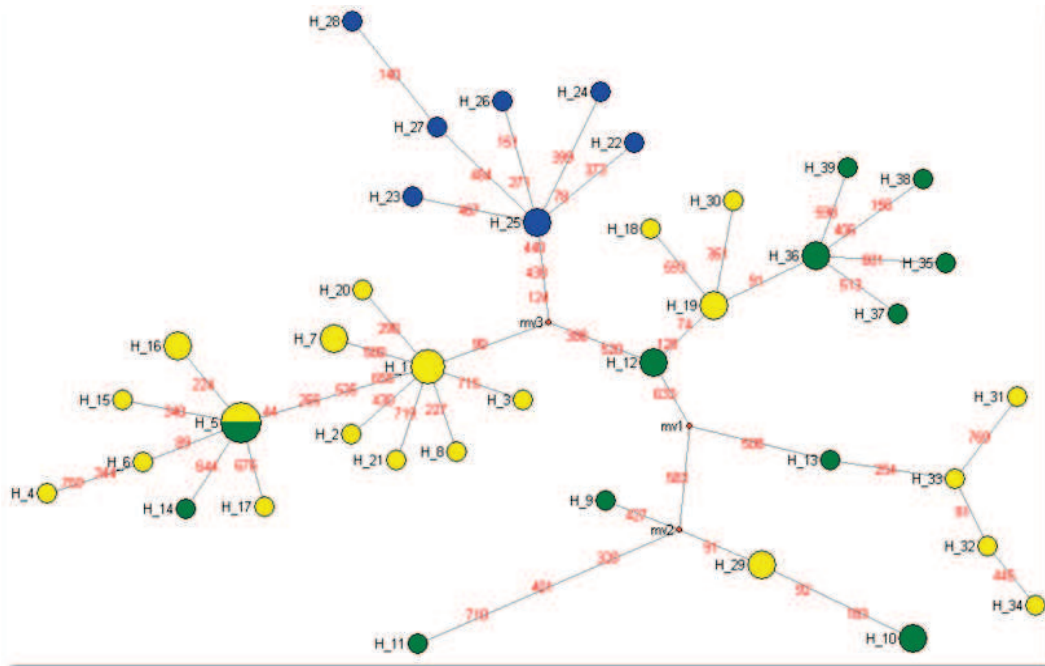


Figure IV.17. Réseau haplotypique de *SAD2_3'* obtenu par la méthode de "median-joining" (Bandelt *et al.*, 1999) grâce au logiciel NETWORK 4.516 (<http://www.fluxus-engineering.com/>). La position des mutations est indiquée en rouge sur les branches. Chaque haplotype est représenté par un cercle de diamètre proportionnel à sa fréquence. Les haplotypes d'Afrique de l'Ouest sont en jaune, les haplotypes d'Afrique Centrale sont en vert et les haplotypes d'Afrique de l'Est sont en bleu. Les points rouges représentent des haplotypes de jonction manquants.

3.2.6 Analyse phylogénétique des gènes *SAD1* et *SAD2*

L'analyse phylogénétique de *SAD1_5'* (**Figure IV.18**) révèle deux grands clades correspondants aux deux groupes Est (*ssp. nilotica*) et Ouest (*ssp. paradoxa*). Cette divergence dans l'arbre est soutenue par une valeur forte de bootstrap (83%) mais qui reste non significative au seuil 5%. On note au sein du clade Ouest une évolution plus marquée entre les séquences avec des longueurs de branche plus longues qu'au sein du clade Est (**Figure IV.18**). L'analyse phylogénétique de *SAD1_5'* (**Figure IV.18**) révèle deux grands clades correspondants aux deux groupes Est (*ssp. nilotica*) et Ouest (*ssp. paradoxa*). Cette divergence dans l'arbre est soutenue par une valeur forte de bootstrap (83%) mais qui reste non significative au seuil 5%. On note au sein du clade Ouest une évolution plus marquée entre les séquences avec des longueurs de branche plus longues qu'au sein du clade Est (**Figure IV.18**).

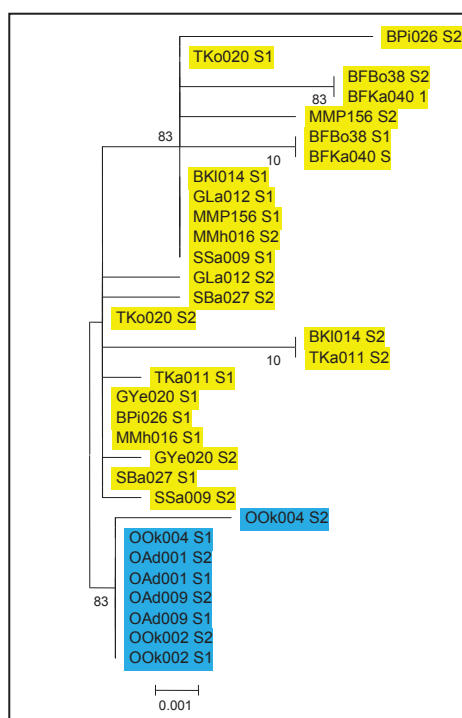


Figure IV.18. Analyse phylogénétique des séquences *SAD1_5'* par Maximum de Vraisemblance. L'arbre a été réalisé avec le logiciel MEGA 5 (Tamura *et al.*, 2007) avec 100000 bootstraps. Les séquences d'origine d'Afrique de l'Ouest sont en jaune et les séquences d'Afrique de l'Est sont en bleu.

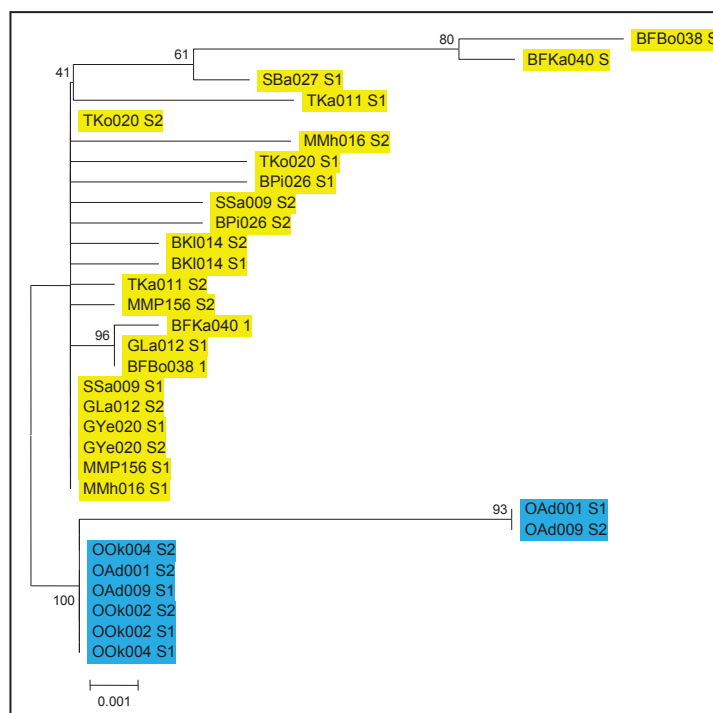


Figure IV.19. Analyse phylogénétique des séquences *SAD2_5'* par Maximum de Vraisemblance. L'arbre a été réalisé avec le logiciel MEGA 5 (Tamura *et al.*, 2007) avec 100000 bootstraps. Les séquences d'origine d'Afrique de l'Ouest sont en jaune et les séquences d'Afrique de l'Est sont en bleu.

L'arbre obtenu par maximum de vraisemblance pour les séquences de *SAD2_5'* est présenté à la **Figure IV.19**. On observe globalement que la topologie de l'arbre est similaire à



Chapitre IV : Étude du polymorphisme des gènes candidats

celle de l'arbre *SAD1_5'*, avec deux grand clades distinguant les deux sous-espèces. En revanche, les branches de l'arbre *SAD2_5'* sont plus longues, témoins d'une accumulation de plus de mutation entre les taxons (séquences).

De la même façon que pour les arbres *SAD1_5'* (**Figure IV.18**) et *SAD2_5'* (**Figure IV.19**), les arbres obtenus par maximum de vraisemblance pour les séquences *SAD1_3'* (**Figure IV.20**) et *SAD2_3'* (**Figure IV.21**) distinguent clairement les séquences de la sous-espèce *nilotica* des séquences de la sous-espèce *paradoxa*, avec la constitution de deux clades. De plus, la comparaison de la topologie des arbres, montre que les branches de l'arbre *SAD1_3'* (**Figure IV.20**) sont plus courtes que celles de l'arbre *SAD2_3'* (**Figure IV.21**). Ceci confirme les résultats de l'analyse du polymorphisme des haplotypes avec les séquences *SAD2_3'* qui ont accumulé plus de mutations que les séquences *SAD1_3'*. Enfin, confirmant les observations des réseaux haplotypiques, on observe que les séquences d'origine d'Afrique de l'Ouest et d'Afrique Centrale sont mélangées au sein du clade « *paradoxa* » des deux arbres *SAD1_3'* (**Figure IV.20**) et *SAD2_3'* (**Figure IV.21**).

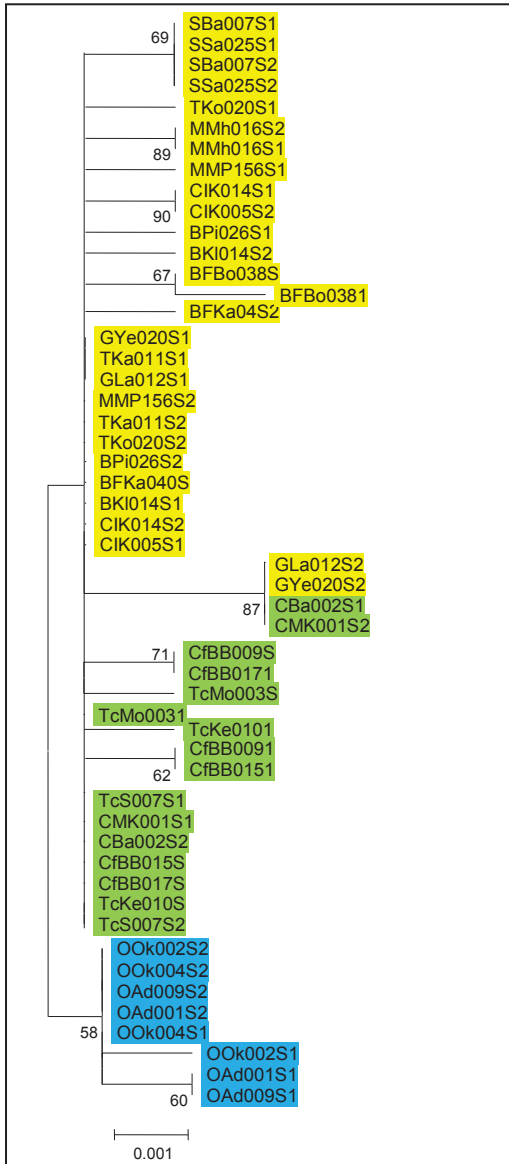


Figure IV.20. Analyse phylogénétique des séquences *SAD1_3'* par Maximum de Vraisemblance. L'arbre a été réalisé avec le logiciel MEGA 5 (Tamura et al., 2007) avec 100000 bootstraps. Les séquences d'origine d'Afrique de l'Ouest sont en jaunes, les séquences d'Afrique Centrale sont en vert et les séquences d'Afrique de l'Est sont en bleu.

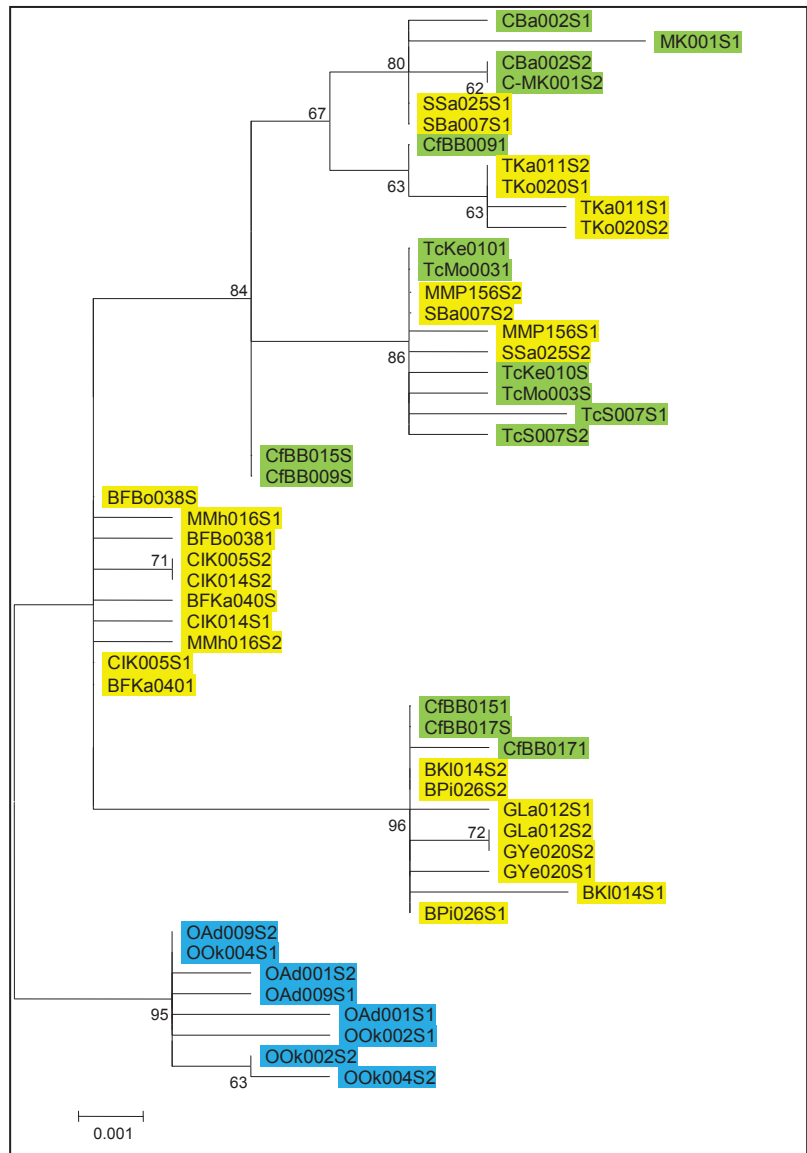


Figure IV.21. Analyse phylogénétique des séquences *SAD2_3'* par Maximum de Vraisemblance. L'arbre a été réalisé avec le logiciel MEGA 5 (Tamura et al., 2007) avec 100000 bootstraps. Les séquences d'origine d'Afrique de l'Ouest sont en jaunes, les séquences d'Afrique Centrale sont en vert et les séquences d'Afrique de l'Est sont en bleu.

3.3 Étude du gène VTE2-2 chez *Vitellaria paradoxa*

3.3.1 Structure du gène VTE2-2 chez le Karité

L'amplification du gène *VTE2-2* chez *Vitellaria paradoxa*, a posé quelques difficultés techniques liées à une disponibilité faible de séquences chez les espèces apparentées et à une conservation médiocre du gène au sein du règne végétal. Ceci a donc rendu plus difficile l'obtention de séquences du gène pleine longueur chez *Vitellaria paradoxa*, et nous nous sommes donc concentrés sur l'amplification et le séquençage d'un fragment de 800bp comprenant l'intron 2, l'exon 3, l'intron 3 et l'exon 4 du gène (**Figure IV.22**). Nous présentons ici les analyses menées sur deux types d'échantillonnages, l'un à l'échelle de l'aire naturelle, et l'autre à l'échelle régionale (Mali) avec 4 populations distribuées selon un gradient climatique Nord-Sud.



Figure IV.22. Structure et région étudiée au sein du gène *VTE2-2*. La flèche noire délimite la partie étudiée dans les analyses suivantes. L'échelle graduée au dessus du gène est exprimée en kb.

3.3.2 Étude du gène VTE2-2 chez *Vitellaria paradoxa* à l'échelle de l'aire naturelle

➤ Polymorphisme du gène VTE2-2 sur l'aire naturelle

Le **Tableau IV.6** présente l'ensemble des résultats obtenus pour l'analyse du polymorphisme des séquences du gène *VTE2-2* à l'échelle de l'aire naturelle. Les résultats montrent dans l'échantillonnage total que le gène *VTE2-2* (**Tableau IV.6**) est moins polymorphe que les gènes *SAD1* et *SAD2* (**Tableau IV.4**). Ainsi, sur 88 allèles séquencés sur 770 paires de bases, seulement 10 sites en ségrégation ont été détectés, résultant en 15 haplotypes avec une diversité haplotypique moyenne de $Hd_{VTE2-2} = 0.273$ (**Tableau IV.6**). Les résultats révèlent également des valeurs très faibles pour les estimateurs π (0.0007) et θ_w (0.00289). Ce résultat confirme ainsi la faible diversité des séquences de *VTE2-2* sur l'aire



Chapitre IV : Étude du polymorphisme des gènes candidats

naturelle. D'autre part, on constate que les exons du gène *VTE2-2* présentent des indices de diversité (H_d , π et θ_w) deux à trois fois inférieurs à ceux observés dans les introns (**Tableau IV.6**). On observe dans l'échantillon d'Afrique de l'Ouest (**Tableau IV.6**), que le gène *VTE2-2* montre un niveau de diversité proche de celui observé au sein de l'aire naturelle avec 8 sites en ségrégation, 11 haplotypes, une diversité haplotypique de 0.295, un $\pi = 0.0008$ et un $\theta_w = 0.0023$. Cependant, cette observation tient grandement au fait que les séquences d'Afrique de l'Ouest ($N_{seq_{OUEST}} = 62$) constituent 75% de l'échantillon total de l'aire naturelle ($N_{seq_{TOTAL}} = 88$) (**Tableau IV.6**).

Au niveau de l'échantillon d'Afrique centrale (**Tableau IV.6**), on observe un niveau de diversité faible de *VTE2-2* avec seulement 3 sites en ségrégation détectés résultant en 4 haplotypes, une diversité haplotypique de 0.216, un $\pi = 0.0004$ et un $\theta_w = 0.0008$. Enfin le niveau de diversité observé au sein de *VTE2-2* pour l'échantillon « Est » est encore plus faible, avec un seul site en ségrégation résultant en 2 haplotypes avec une diversité moyenne de $H_d = 0.25$, un $\pi = 0.0003$ et un $\theta_w = 0.0005$. Cependant, il convient de noter que le groupe Afrique de l'Est n'est représenté que par 8 séquences contre 62 en Afrique de l'Ouest et 18 en Afrique Central.

- Étude du déséquilibre de liaison (DL) au sein du gène *VTE2-2* sur l'échantillon de l'aire naturelle

L'analyse du DL au sein du gène *VTE2-2* montre une valeur globale de R^2 faible et non significative au regard du test d'exactitude de Fisher (**Figure IV.23**). On note cependant quelques valeurs significatives pour certaines paires de nucléotides testées au sein de l'échantillon total, mais également au sein de l'échantillon Ouest (**Figure IV.23**). En revanche, l'analyse de DL n'a pu être déterminée au sein des échantillons d'Afrique Centrale et d'Afrique de l'Est du fait de l'absence de polymorphisme significatif au sein de ces échantillons pour le gène *VTE2-2*.

Chapitre IV : Étude du polymorphisme des gènes candidats

Tableau IV.6. Diversité et sélection au sein du gène *VTE2-2* considérant différents échantillons : Total aire naturelle = les séquences de l'aire naturelle ; Ouest = séquences des individus d'Afrique de l'Ouest ; Centre = séquences d'Afrique Centrale ; Est = séquences d'Afrique de l'Est. N seq : nombre de séquences ; N site : nombre de nucléotides ; S : nombre de sites en ségrégation ; Hap : nombre d'haplotypes ; Hd : diversité haplotypique (Nei, 1987) ; π : estimateur de la diversité nucléotidique de Nei (Nei, 1987) ; θ : estimateur de la diversité nucléotidique de Watterson (Watterson, 1975) ; D : valeur du D de Tajima (Tajima, 1989) ; D* et F* : valeur des D* et F* de Fu et Li (Fu & Li, 1993). # : P<0.05 ; * : P< 0.01.

Gène		VTE2-2		
Populations	Domain Région	Introns [1↔115] [269↔652]	Exons [116↔268] [653↔770]	Total 1-770
Total Aire naturelle	N seq	88	88	88
	N sites	499	271	770
	S	9	1	10
	Hap	11	3	15
	Hd	0.273	0.109	0.273
	π	0.00087	0.0004	0.0007
	θ_w	0.00369	0.00146	0.00289
	D	-1.9444 *	-1.1405	-2.0114 *
	D*	-2.3506 *	-1.0405	-2.4580 *
	F*	-2.6194 *	-1.2509	-2.7324 *
Ouest	N seq	62	62	62
	N sites	499	271	770
	S	7	1	8
	Hap	10	2	11
	Hd	0.295	0.123	0.295
	π	0.00102	0.00045	0.00082
	θ_w	0.00308	0.00079	0.00226
	D	-1.6987 *	-0.5397	-1.6737 #
	D*	-2.0134 #	0.5256	-1.6886 #
	F*	-2.2491 #	0.2484	-1.9846 #
Centre	N seq	18	18	18
	N sites	499	271	770
	S	2	1	3
	Hap	3	2	4
	Hd	0.209	0.111	0.216
	π	0.00042	0.00041	0.00042
	θ_w	0.00059	0.00107	0.00076
	D	-0.529	-1.1647	-1.0963
	D*	0.6669	-1.4995	-0.5522
	F*	0.4045	-1.6117	-0.7978
Est	N seq	8	8	8
	N sites	499	271	770
	S	1	0	1
	Hap	2	1	2
	Hd	0.25	0	0.25
	π	0.0005	0	0.00033
	θ_w	0.00078	-	0.0005
	D	-1.1264	-	-1.1264
	D*	-1.2035	-	-1.2035
	F*	-	-	-

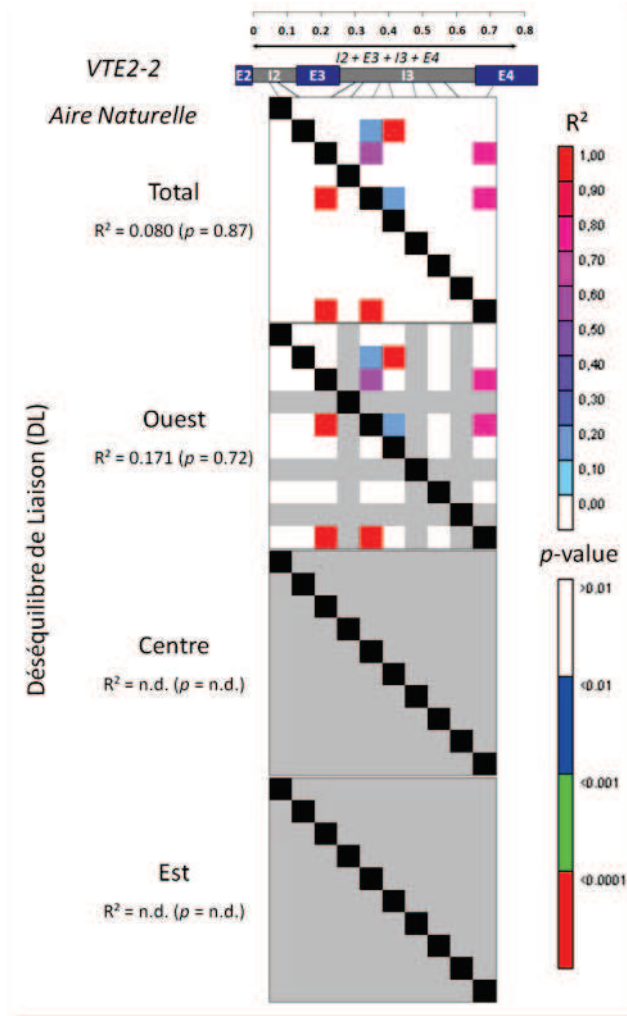


Figure IV.23. Déséquilibre de liaison au sein du gène *VTE2-2* considérant différents échantillons : Total = toutes les séquences disponibles de l'aire naturelle ; Ouest = séquences des individus d'Afrique de l'Ouest ; Centre = séquences d'Afrique Centrale ; Est = séquences d'Afrique de l'Est). Les matrices triangulaires supérieures indiquent les valeurs des R^2 entre paires de nucléotides. Les matrices triangulaires inférieures indiquent les degrés de significativité des valeurs de DL obtenu avec le test d'exactitude de Fisher.

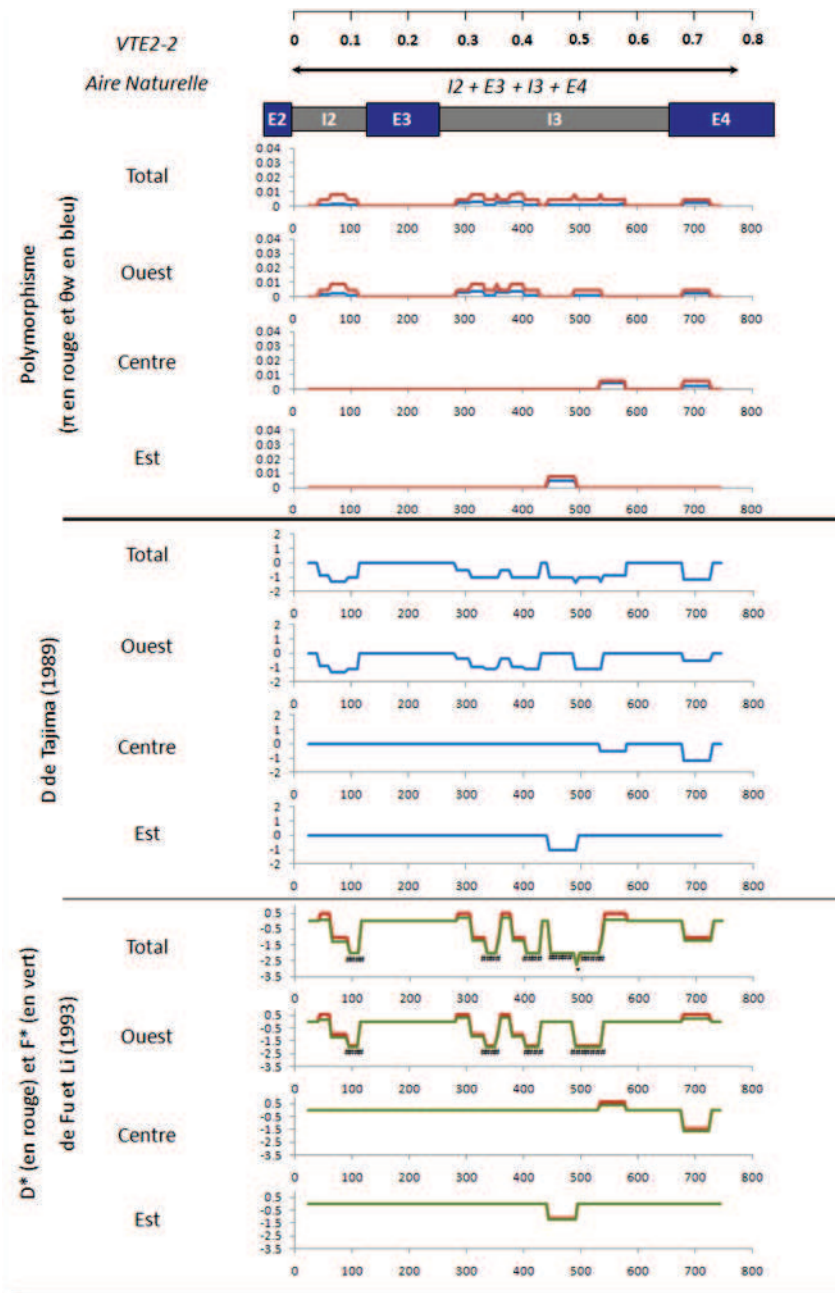


Figure IV.24. Évolution de la diversité et des indices de sélection du gène *VTE2-2* à l'échelle de l'aire naturelle, en utilisant la méthode de la « sliding window » du logiciel DNAsP v. 5.10 (Librado & Rozas, 2009) au sein de différents échantillons : Total = séquences de l'aire naturelle ; Ouest = séquences d'Afrique de l'Ouest ; Centre = séquences d'Afrique Centrale ; Est = séquences d'Afrique de l'Est. π : estimateur de la diversité nucléotidique de Nei (Nei, 1987); θ_w : estimateur de la diversité nucléotidique de Watterson (Watterson, 1975); D : valeur du D de Tajima (Tajima, 1989) ; D* et F* : valeur des D* et F* de Fu et Li (Fu & Li, 1993). # : $P < 0.05$; * : $P < 0.01$.



➤ Étude de la sélection du gène *VTE2-2* sur l'aire naturelle

Sur l'échantillon total, le test de Tajima montre que les séquences *VTE2-2* présentent un excès significatif de sites polymorphes rares avec $D = -2.01$ ($p < 0.02$), et les tests de Fu et Li montrent un excès significatif de sites polymorphes récents avec $D^* = -2.46$ ($p < 0.02$) et $F^* = -2.73$ ($p < 0.02$) (**Tableau IV.6**). Cette observation reste vraie également au sein de l'échantillon d'Afrique de l'Ouest avec des tests de Tajima ($D = -1.67$; $p < 0.05$) et de Fu et Li ($D^* = -1.69$, $p < 0.05$; $F^* = -1.98$, $p < 0.05$) significativement négatifs (**Tableau IV.6**). En revanche, au sein des échantillons « Est » et « Centre », les tests de neutralité ne sont plus significatifs (**Tableau IV.6**), suggérant donc un possible impact contrasté de la sélection sur l'aire naturelle.

L'évolution des valeurs de D , D^* et F^* le long du fragment étudié de *VTE2-2* obtenu par la méthode de « sliding window » est présentée à la **Figure IV.24**. Ainsi, bien que le test de Tajima soit globalement significativement négatif dans l'échantillon total (**Tableau IV.6**), la méthode de « sliding window » ne permet pas de mettre en évidence une région du gène significativement plus sélectionnée. On note cependant que les tests de Fu et Li sont très significativement négatifs au sein des deux introns du gène pour l'échantillon global, mais également pour le sous-échantillon d'Afrique de l'Ouest. À l'inverse, confirmant les résultats précédents, aucune valeur significative de test n'a pu être détectée au sein de *VTE2-2* que ce soit pour l'échantillon « Est » ou « Centre » (**Figure IV.24**). Ainsi, comme nous l'avons également constaté lors de l'analyse des gènes *SAD1* et *SAD2*, les tests de neutralité ne sont pas homogènes le long du gène avec globalement plus de valeurs significatives au sein des introns (**Tableau IV.4 et Figure IV.24**).

➤ Différenciation génétique et réseau d'haplotypes du gène *VTE2-2* sur l'aire naturelle

Afin de quantifier le niveau de différenciation génétique entre les populations au sein du gène *VTE2-2*, nous avons calculé les indices de F_{ST} (Hudson *et al.*, 1992) et de G_{ST} (Nei, 1973) à l'aide de DnaSP v. 5.10 (Rozas and Rozas, 1999).

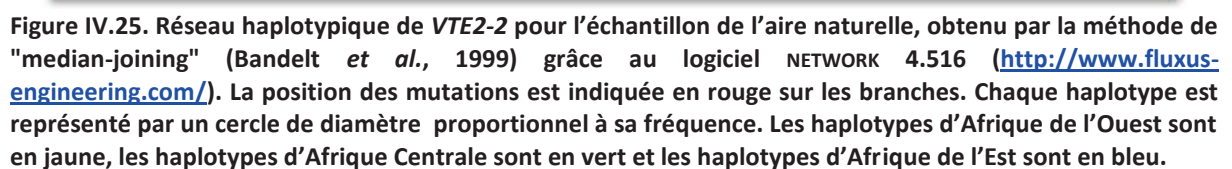
Ainsi, le **Tableau IV.7** présente les indices de différenciation génétique entre les populations d'Afrique de l'Ouest, d'Afrique Centrale et d'Afrique de l'Est pour le fragment

analysé du gène *VTE2-2*. On observe des valeurs de F_{ST} et de G_{ST} très faibles avec de valeurs de $F_{ST} < 0.05$ et de $G_{ST} < 0.03$ (**Tableau IV.7**). Ces valeurs sont en outre plus faibles que celle habituellement constatées chez *Vitellaria paradoxa*, et en particulier celles observées sur les gènes *SAD1* et *SAD2*. On note cependant que la différenciation est légèrement plus faible entre les groupes « Centre » et « Est » qu'entre les groupes « Centre » et « Ouest », confirmant ainsi les résultats observés dans le Chapitre II avec les marqueurs neutres (microsatellites nucléaires).

Tableau IV.7. Niveaux de différenciation génétique F_{ST} (en noir) et G_{ST} (en rouge) du gène *VTE2-2* entre les populations Ouest (Afrique de l'Ouest), Centre (Afrique Centrale) et Est (Afrique de l'Est). Les valeurs sont significatives au seuil 5%. n.m. : valeur non mesurée.

VTE2-2	Population	Ouest	Centre	Est
	Ouest	-	0.042	0.043
	Centre	0.012	-	0.019
	Est	0.025	0.009	-

Le réseau haplotypique (**Figure IV.25**) reconstruit par la méthode du « median-joining » (Bandelt *et al.*, 1999), montre les relations phylogénétiques entre les haplotypes du gène *VTE2-2* sur l'échantillon de l'aire naturelle. Ainsi, on observe un haplotype très majoritaire (H_1) partagé par les trois populations de l'aire naturelle (**Figure IV.25**). Cet haplotype semble également à l'origine directe d'au moins 9 autres haplotypes détectés (H_3, H_4, H_5, H_7, H_8, H_10, H_12, H_13 et H_15). Il pourrait donc constituer l'haplotype ancestral du gène. On remarque également que les haplotypes spécifiques d'Afrique de l'Est et d'Afrique centrale présentent une faible divergence avec l'haplotype H_1 supposé ancestral (**Figure IV.25**). En revanche, on note que trois haplotypes d'Afrique de l'Ouest (H_2, H_6 et H_9) montrent une divergence très importante par rapport aux autres (au minimum 7pb). Hormis ces trois haplotypes d'Afrique de l'Ouest, on constate que les autres allèles du gène *VTE2-2* semblent très proches, étant distants les uns des autres de seulement un point de mutation (**Figure IV.25**).



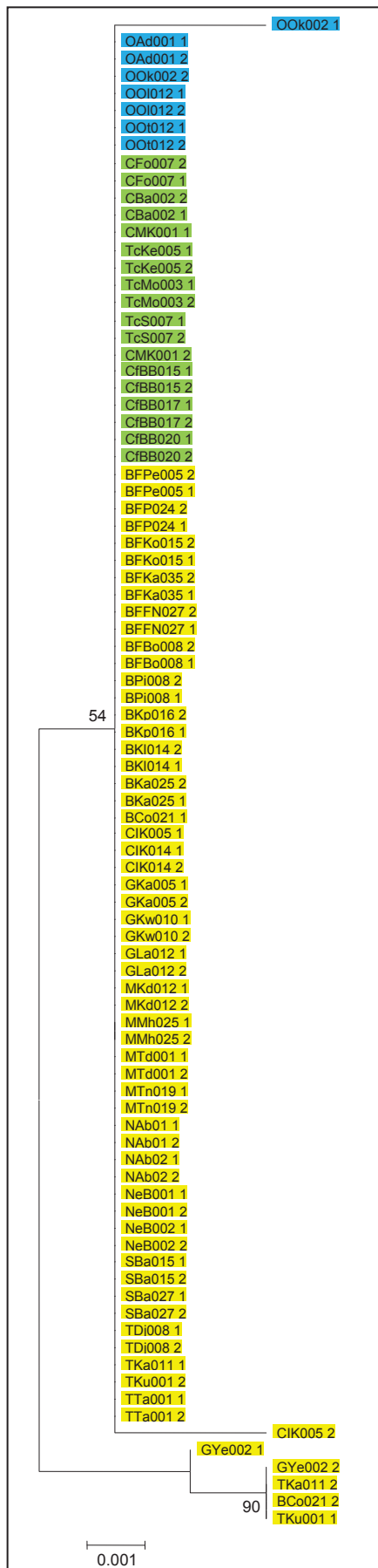


Figure IV.26. Analyse phylogénétique des séquences de VTE2-2 à l'échelle de l'aire naturelle par Maximum de Vraisemblance. L'arbre a été réalisé avec le logiciel MEGA5 (Tamura *et al.*, 2007) avec 100000 bootstraps.



➤ Analyse phylogénétique du gène VTE2-2 sur l'échantillon de l'aire naturelle

L'étude phylogénétique des séquences du gène *VTE2-2* (**Figure IV.26**) confirme la faible différenciation observée entre les groupes « Ouest », « Centre » et « Est ». Ainsi, l'arbre obtenu par la méthode du maximum de vraisemblance et présenté à la **Figure IV.26**, révèle deux clades ne permettant qu'une très mauvaise discrimination des séquences échantillonnées. En outre, les séquences des différentes origines (Afrique de l'Ouest, Afrique Centrale et Afrique de l'Est) apparaissent mélangées au sein d'un grand clade (**Figure IV.26**). Seules quelques séquences d'Afrique de l'Ouest se distinguent des autres au sein d'un petit clade soutenu à 90%. De plus, en raison du faible niveau de polymorphisme mis en évidence au sein du gène *VTE2-2* dans l'échantillon total de l'aire naturelle, l'arbre montre des longueurs de branches très courtes entre les séquences étudiées. Ainsi les relations phylogénétiques entre les séquences de *VTE2-2* sont moins claires que celles observées entre les séquences du gène *SAD2* notamment (**Figures IV.20 et IV.21**). Contrairement à ce que nous avons observé pour les séquences des gènes *SAD1* et *SAD2*, l'analyse phylogénétique des séquences de *VTE2-2*, ne permet de distinguer ni les provenances géographiques, ni les deux sous-espèces de karité, *paradoxa* et *nilotica*.

3.3.3 Étude du gène VTE2-2 chez *Vitellaria paradoxa* du gradient Nord-Sud au Mali

➤ Étude du polymorphisme

L'analyse du polymorphisme des séquences de *VTE2-2* au sein du gradient Nord-Sud au Mali est présentée dans le Tableau IV.4. Ainsi, les résultats obtenus montrent dans l'échantillonnage « Mali » que le niveau de diversité détecté au sein du gène *VTE2-2* (**Tableau IV.8**) est assez faible, avec sur 76 séquences d'allèles analysés 14 sites en ségrégation, résultant en 12 haplotypes de diversité moyenne $Hd = 0.451$, une valeur de l'estimateur $\pi = 0.0017$ et une valeur de l'estimateur $\theta_w = 0.0041$ (**Tableau IV.8**). Néanmoins, ce niveau de diversité observé au sein du gradient Mali est très supérieur à celui mis en évidence sur l'aire naturelle ($\pi = 0.0007$ et $\theta_w = 0.00289$). En revanche, de façon similaire à ce qui a été observé au sein de l'échantillon « aire naturelle », les exons du gène *VTE2-2* présentent des indices de diversité (Hd , π et θ_w) très inférieurs à ceux observés dans les introns (**Tableau IV.8**).

Chapitre IV : Étude du polymorphisme des gènes candidats

Tableau IV.8. Diversité et sélection au sein du gène *VTE2-2* considérant différents échantillons : Mali = total des séquences du gradient ; Na = séquences de Nafégué ; Mp = séquences de Mperesso ; Da = séquences de Daelan ; Sa = séquences de Sassambourou. N seq : nombre de séquences ; N site : nombre de nucléotides ; S : nombre de sites en ségrégation ; Hap : nombre d'haplotypes ; Hd : diversité haplotypique (Nei, 1987) ; π : estimateur de la diversité nucléotidique de Nei (Nei, 1987) ; θ_w : estimateur de la diversité nucléotidique de Watterson (Watterson, 1975) ; D : valeur du D de Tajima (Tajima, 1989) ; D* et F* : valeurs des D* et F* de Fu et Li (Fu & Li, 1993). # : $P < 0.05$; * : $P < 0.01$.

Populations	Gène	VTE2-2		
	Domain Région	Introns [1↔115] [269↔652]	Exons [116↔268] [653↔770]	Total 1-770
Mali	N seq	76	76	76
	N sites	499	271	770
	S	13	1	14
	Hap	14	2	15
	Hd	0.451	0.026	0.451
	π	0.00263	0.0001	0.00172
	θ_w	0.0059	0.00075	0.00405
	D	-1.5748 #	-1.0583	-1.6539 #
	D*	-1.6633	-1.9588 #	-1.9914 #
	F*	-1.9390 #	-1.9658 #	-2.2249 #
Na	N seq	20	20	20
	N sites	499	271	770
	S	6	0	6
	Hap	7	1	7
	Hd	0.505	0	0.505
	π	0.00288	0	0.00185
	θ_w	0.00345	-	0.00222
	D	-0.5289	-	-0.5289
	D*	-0.1542	-	-0.1542
	F*	-0.2999	-	-0.2999
Mp	N seq	22	22	22
	N sites	499	271	770
	S	2	0	2
	Hap	3	1	3
	Hd	0.173	0	0.173
	π	0.00035	0	0.00023
	θ_w	0.00055	-	0.00036
	D	-0.6411	-	-0.6411
	D*	0.635	-	0.635
	F*	0.3364	-	0.3364
Da	N seq	20	20	20
	N sites	499	271	770
	S	11	1	12
	Hap	9	2	9
	Hd	0.732	0.1	0.732
	π	0.00505	0.00037	0.00337
	θ_w	0.00699	0.00104	0.00485
	D	-0.9979	-1.1644	-1.1109
	D*	-1.4114	-1.5396	-1.592
	N seq	-1.4981	-1.6477	-1.6851
Sa	N seq	14	14	14
	N sites	499	271	770
	S	6	0	6
	Hap	3	1	3
	Hd	0.275	0	0.275
	π	0.00175	0	0.00113
	θ_w	0.00386	-	0.00248
	D	-1.9589 *	-	-1.9589 *
	D*	-2.5156 *	-	-2.5156 *
	F*	-2.7036 *	-	-2.7036 *



Chapitre IV : Étude du polymorphisme des gènes candidats

On observe que les niveaux de diversité ne sont pas homogènes au sein du gradient Nord-Sud au Mali. Ainsi, le sous-échantillon du site de Daelan montre le niveau de diversité le plus élevé, avec sur 20 séquences analysées un total de 12 sites en ségrégation, 9 haplotypes, une diversité haplotypique de 0.732, un $\pi = 0.0033$ et un $\theta_w = 0.0048$. Le site de Nafégué, à l'extrême sud du gradient Malien montre un niveau de diversité intermédiaire avec 6 sites polymorphes, 7 haplotypes, une diversité haplotypique de $H_d = 0.505$, un $\pi = 0.0018$ et un $\theta_w = 0.0022$. Le site de Sassambourou, à l'extrême nord, révèle un polymorphisme plus faible avec 6 sites polymorphes, 3 haplotypes, une diversité haplotypique de $H_d = 0.275$, un $\pi = 0.0011$ et un $\theta_w = 0.0025$. Enfin, le site de Mperesso révèle le niveau de diversité nucléotidique très faible avec pour 22 séquences de *VTE2-2* analysées, seulement 2 sites polymorphes (soit 3 haplotypes), une diversité haplotypique de $H_d = 0.173$, un $\pi = 0.0002$ et un $\theta_w = 0.0004$.

➤ Étude du déséquilibre de liaison (DL) au sein du gène *VTE2-2* sur l'échantillon Mali

L'analyse du DL au sein du gène *VTE2-2* pour l'échantillon global du gradient Nord-Sud au Mali, montre une valeur globale de R^2 faible et non significative au regard du test d'exactitude de Fisher (**Figure IV.27**). Ce calcul de déséquilibre de liaison porte sur un nombre réduit de comparaisons par paire de nucléotides polymorphes. Seules quelques unes de ces comparaisons sont significatives selon le test d'exactitude de Fisher (**Figure IV.27**). On note cependant que le déséquilibre de liaison n'est pas homogène, que ce soit en fonction des échantillons testés, ou selon la région du gène analysé. En effet, le nombre de paires de nucléotide en déséquilibre de liaison significatif est nul pour le site de Mperesso, faible pour le site de Nafégué, et plus important pour les sites de Daelan et Sassambourou (**Figure IV.27**). On note pour les sites de Daelan et surtout de Sassambourou, la présence d'un bloc de nucléotides (« haplobloc ») situé autour de l'exon 3 du gène *VTE2-2*, où le déséquilibre de liaison est fort (**Figure IV.27**).

➤ Étude de la sélection du gène *VTE2-2* sur le gradient Nord-Sud au Mali

Sur l'échantillon global du gradient Nord-Sud Malien, les tests de neutralité suggèrent l'impact d'une sélection directionnelle du gène *VTE2-2* (**Tableau IV.8**). Ainsi, le test de Tajima montre que les séquences *VTE2-2* du gradient Malien présentent un excès significatif de

sites polymorphes rares avec $D = -1.6539$ ($p < 0.05$), et les tests de Fu et Li indiquent un excès significatif de sites polymorphes récents avec $D^* = -1.9914$ ($p < 0.05$) et $F^* = -2.2249$ ($p < 0.05$) (**Tableau IV.8**).

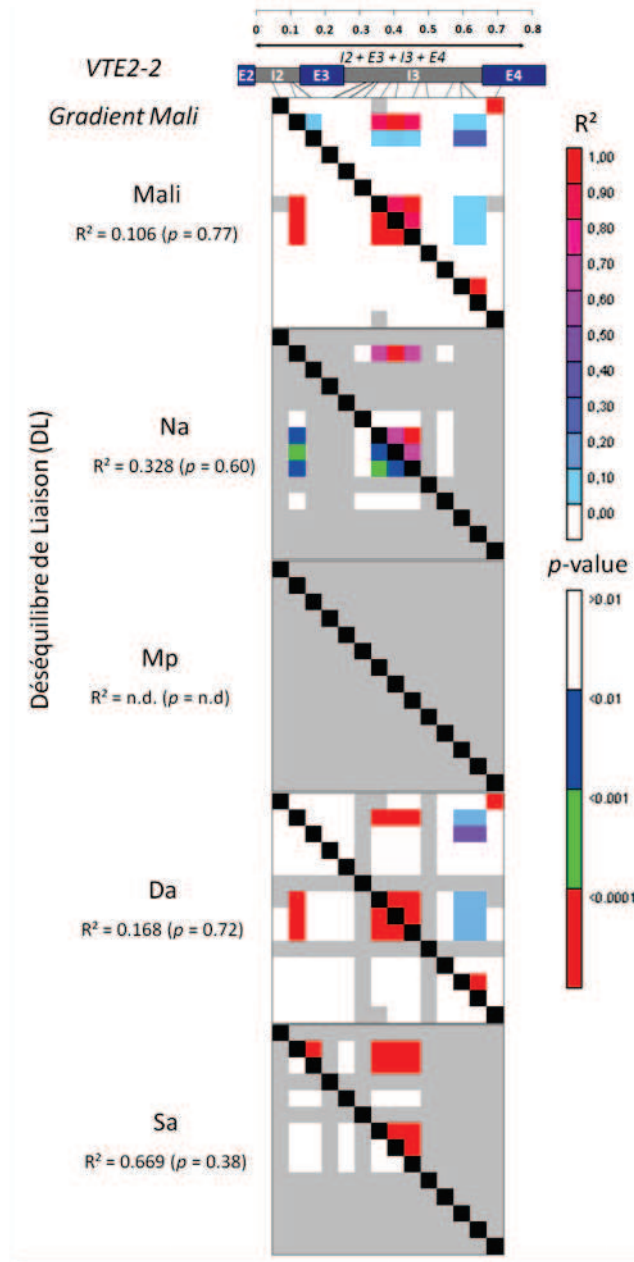


Figure IV.27. Déséquilibre de liaison au sein du gène *VTE2-2* considérant différents échantillons : Mali = toutes les séquences du gradient Nord-Sud ; Na = séquences des individus de Nafégué ; Mp = séquences des individus de Mpéresso ; Da = séquences des individus de Daelan ; Sa = séquences des individus de Sassambourou. Les matrices triangulaires supérieures indiquent les valeurs des R^2 entre paires de nucléotides. Les matrices triangulaires inférieures indiquent les degrés de significativité des valeurs de DL obtenus avec le test d'exactitude de Fisher.

Les tests de neutralité sont non significatifs au sein des sites de Nafégué ($D = -0.5289$ n.s. ; $D^* = -0.1542$ n.s. ; $F^* = -0.2999$ n.s.), Mperesso ($D = -0.6411$ n.s. ; $D^* = 0.635$ n.s. ; $F^* = 0.3364$ n.s.) et Daelan ($D = -1.1109$ n.s. ; $D^* = -1.592$ n.s. ; $F^* = -1.6851$ n.s.). En



Chapitre IV : Étude du polymorphisme des gènes candidats

revanche, on note que les tests de sélection sont négatifs et très significatifs au sein de l'échantillon « Sassambourou », avec un D de Tajima = -1.9589 ($p < 0.02$), D^* de Fu et Li = -2.5156 ($p < 0.02$) et un F^* de Fu et Li = -2.7036 ($p < 0.05$). Ce résultat suggère ainsi une sélection locale directionnelle au Nord du Mali au sein du site de Sassambourou (**Tableau IV.8**).

Au sein du gradient Nord-Sud Malien, l'évolution des valeurs de D , D^* et F^* le long du fragment étudié de *VTE2-2* obtenu par la méthode de « sliding window » est présentée à la **Figure IV.28**. Ainsi, de même que ce qui avait été constaté au sein de l'échantillonnage « aire naturelle », la **Figure IV.28** révèle que les tests de neutralités ne sont pas homogènes le long du gène avec plus de valeurs négatives et significatives au sein des introns qu'au sein des exons. On constate cependant des valeurs significativement négatives au seuil 5%, au sein de l'exon 4 du gène en considérant l'échantillon global du Mali.

On note de nombreuses valeurs significativement négatives pour les tests de Tajima et de Fu et Li dans les introns 2 et 3 du gène *VTE2-2*, considérant les séquences d'origine de Sassambourou au nord du gradient Mali (**Figure IV.28**). Ce résultat suggère donc la présence de régions positivement sélectionnées au voisinage de ces introns. Nous faisons également la même observation pour l'échantillon de Daelan qui présente pourtant des tests de sélection globalement non significatifs (**Tableau IV.8**). En revanche, aucune valeur significative n'a été détectée au sein des échantillons de séquences de Nafégué et Mpéresso (les deux sites les plus au sud du gradient), confirmant les résultats obtenus **Tableau IV.8**.

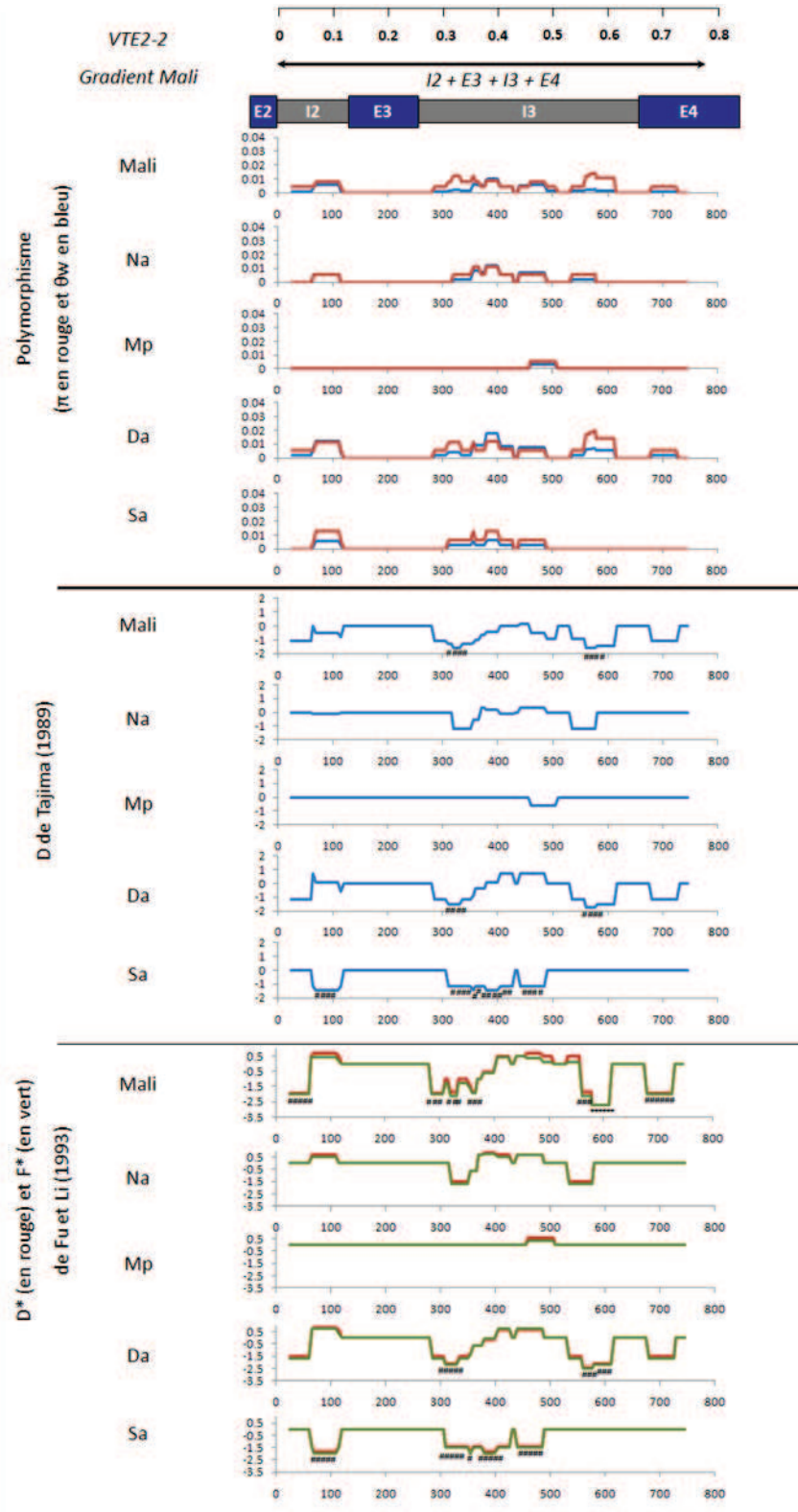


Figure IV.28. Évolution de la diversité et des indices de sélection du gène *VTE2-2* au sein du gradient Nord-Sud Malien, en utilisant la méthode de la « sliding window » du logiciel DNASP v. 5.10 (Librado & Rozas, 2009). Différents échantillons sont considérés : Mali = toutes les séquences du gradient Nord-Sud ; Na = séquences des individus de Nafégué ; Mp = séquences des individus de Mpéresso ; Da = séquences des individus de Daelan ; Sa = séquences des individus de Sassambourou. π : estimateur de la diversité nucléotidique de Nei (Nei, 1987). θ_w : estimateur de la diversité nucléotidique de Watterson (Watterson, 1975); D : valeur du D de Tajima (Tajima, 1989) ; D* et F* : valeur des D* et F* de Fu et Li (Fu & Li, 1993). # : $P < 0.05$; * : $P < 0.01$.



- Différenciation génétique et réseau d'haplotypes du gène VTE2-2 sur l'aire naturelle

On constate que les niveaux de différenciation génétiques entre les sites du gradient Nord-Sud au Mali sont très variables. Ainsi, on observe (**Tableau IV.9**) des valeurs très faibles de différenciation entre les sites de Nafégué, Daelan et Sassambourou avec notamment des valeurs de $F_{ST} < 0.07$ (Hudson *et al.*, 1992). En revanche, la différenciation observée entre le site de Mperesso et les sites de Nafégué et de Daelan est plus importante avec des $F_{ST} > 0.12$ (**Tableau IV.9**). Cependant ces niveaux de différenciation restent faibles, et les valeurs de G_{ST} (Nei, 1973) confirment ce résultat avec des valeurs systématiquement inférieures à 0.07. Ces niveaux de différenciation sont similaires à ceux observés dans les études précédentes sur des loci neutres (Sanou *et al.*, 2005).

Tableau IV.9. Niveaux de différenciation génétique F_{ST} (en noir) et G_{ST} (en rouge) du gène VTE2-2 entre les sites du gradient Nord-Sud du Mali. Les valeurs sont significatives au seuil 5%. n.m. : valeur non mesurée.

VTE2-2	Population	Nafégué	Mperesso	Daelan	Sassambourou
	Nafégué	-	0.126	0.016	0.020
	Mperesso	0.027	-	0.187	0.014
	Daelan	0.009	0.054	-	0.063
	Sassambourou	0.058	0.006	0.067	-

On observe à la **Figure IV.29**, le réseau haplotypique du gène VTE2-2 au sein du gradient Mali, reconstruit par la méthode du « median-joining » (Bandelt *et al.*, 1999). On observe en outre l'émergence d'un haplotype (H_3) très majoritaire dans l'échantillon. Cet allèle correspond en réalité à l'haplotype H_1 observé à l'échelle de l'aire naturelle (**Figure IV.25**). Cet haplotype (H_3) est partagé par les individus des quatre sites du gradient Nord-Sud au Mali. De la même façon que pour l'échantillon de l'aire naturelle, H_3 est à l'origine de nombreux autres allèles détectés dans l'échantillon (H_1, H_5, H_6, H_7, H_10, H_11, H_12 et H_14) et pourrait constituer un haplotype ancestral du gène VTE2-2 au sein de l'échantillon Mali (**Figure IV.29**). En revanche, on remarque que 6 haplotypes (H_2, H_4, H_8, H_9, H_13 et H_15) sont très distants de l'haplotype central et concernent les 3 des sites Mali (Daelan, Sassambourou et Nafégué). En particulier, l'haplotype H_8 montre une divergence de 16 nucléotides d'avec l'haplotype ancestral H_3 (**Figure IV.29**).

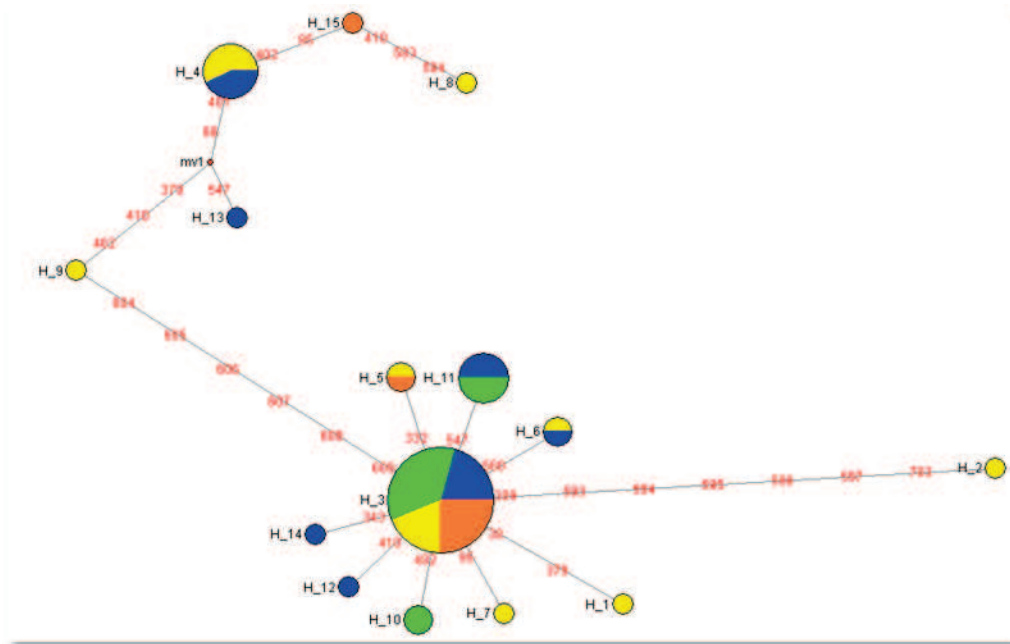


Figure IV.29. Réseau haplotypique de *VTE2-2* pour gradient Nord-Sud Malien, obtenu par la méthode de "median-joining" (Bandelt *et al.*, 1999) grâce au logiciel NETWORK 4.516 (<http://www.fluxus-engineering.com/>). La position des mutations est indiquée en rouge sur les branches. Chaque haplotype est représenté par un cercle de diamètre proportionnel à sa fréquence. Les haplotypes de Nafégué sont en bleu, les haplotypes de Mperesso sont en vert, les haplotypes de Daelan sont en jaune et les haplotypes de Sassambourou sont en orange. Les points rouges représentent des haplotypes de jonction manquant dans l'échantillonnage.

➤ Analyse phylogénétique du gène *VTE2-2* sur le gradient Nord-Sud au Mali

La reconstruction d'un arbre décrivant les relations phylogénétiques entretenues entre les séquences du gène *VTE2-2* obtenues au sein du gradient Nord-Sud échantillonné au Mali, est présentée à la **Figure IV.30**. On observe une topologie d'arbre peu discriminante, avec un grand clade en forme de râteau, soutenu à 80% et concentrant la quasi-totalité des séquences étudiées (**Figure IV.30**). Cet arbre obtenu par la méthode du maximum de vraisemblance, confirme la faible différenciation des sites échantillonnés. Ainsi, de même que pour l'étude phylogénétique à l'échelle de l'aire naturelle, le faible niveau de polymorphisme observé au sein du gène *VTE2-2* dans l'échantillon Mali est responsable des longueurs de branches courtes entre les séquences étudiées (**Figure IV.30**). Ce patron de différenciation est cependant cohérent avec l'hypothèse d'une sélection directionnelle, souvent responsable des topologies d'arbres en « râteau ».

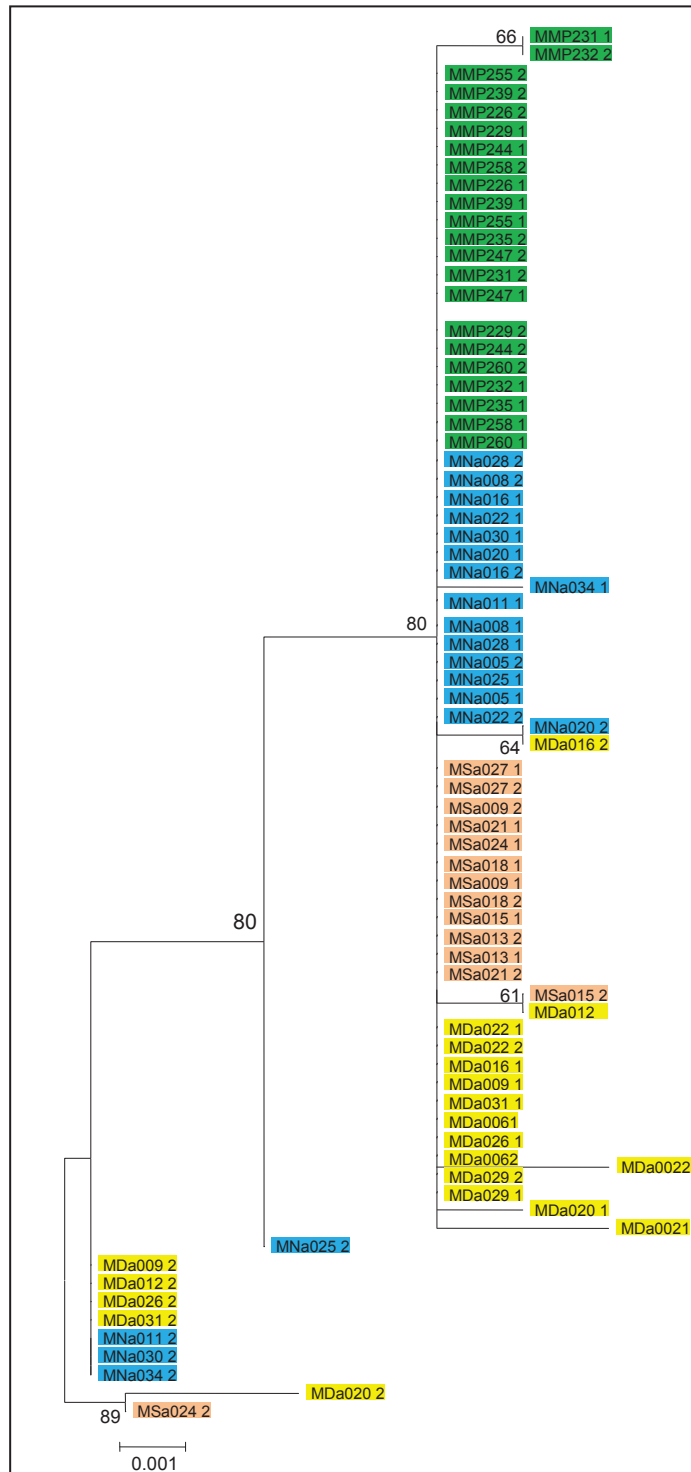


Figure IV.30. Analyse phylogénétique des séquences de *VTE-2* à l'échelle du gradient Nord-Sud au Mali par Maximum de Vraisemblance. L'arbre a été réalisé avec le logiciel MEGA 5 (Tamura et al., 2007) avec 100000 bootstraps.

➤ Structure génétique des sites du gradient Nord-Sud au Mali

Tous les loci microsatellites nucléaires sont polymorphes et possèdent de 4 allèles pour mCIRVp175 à 19 allèles pour mCIRVp168 par locus. Le nombre moyen d'allèles par locus et par population varie peu, ayant une valeur minimale pour la population Mperesso ($A_s = 5,55$) et maximale pour Nafégué ($A_s = 6,27$). Les indices de fixation F_{IS} sont caractérisés par des valeurs significatives et positives indiquant un déficit en hétérozygotes à chaque site étudié (**Tableau IV.10**). L'indice de différenciation génétique est très faible et non significatif ($F_{ST} = 0,005$). Ainsi, l'analyse Bayésienne de la structure génétique des sites étudiés par le logiciel STRUCTURE (Pritchard *et al.*, 2000) ne permet pas de différencier clairement les sites. Cependant, l'examen des log de probabilité par la méthode d'Evanno (Evanno *et al.*, 2005) suggère 4 clusters (**Figure IV.31**).

Tableau IV.10. Paramètres de diversité génétique des sites du gradient Nord-Sud au Mali. Hexp: l'hétérozygotie attendue ; Hobs : l'hétérozygotie observée ; F_{IS} : l'index de fixation ; F_{ST} : l'indice de différenciation génétique. * valeurs significatives à 5%

Populations	N	Hexp	Hobs	F_{IS}	F_{ST}
Sassambourou	30	0.47	0.48	0.138*	0.005
Daelan	30	0.46	0.47	0,120*	
Mperesso	30	0.46	0.47	0.102*	
Nafégué	30	0.47	0.47	0.112*	

➤ Association génotype/phénotype au sein du gradient Nord-Sud au Mali

Les résultats des tests d'association effectués à l'aide du logiciel TASSEL (Bradbury *et al.*, 2007) selon les deux modèles d'association, tenant compte ou non de la structure génétique des individus sont présentés dans le **Tableau IV.11**. Que soit pour les teneurs en α -, β -, γ -, ou δ -tocophérols, aucune association significative n'a pu être mise en évidence avec l'utilisation du modèle linéaire ne tenant pas compte de la structure génétique des individus dans l'échantillonnage (**Tableau IV.11**). Avec l'utilisation du modèle tenant compte de la structure génétique, une seule position nucléotidique (331 sur la séquence) du le fragment du gène *VTE2-2* étudié, a pu être associée significativement à la variation des teneurs en α -tocophérol (**Tableau IV.11**). Ainsi, on note pour cette position nucléotidique située au cœur de l'intron 3 du gène *VTE2-2*, une p -value significative pour le test d'exactitude de Fisher ($F = 4.32$; $p = 0.015$) avec un $R^2 = 0.041$ (ajusté par la structure génétique des individus), signifiant que la variation nucléotidique observée à la position 331 du fragment étudié, explique plus de 4% de la variation globale des teneurs en α -tocophérol

(Tableau IV.11).

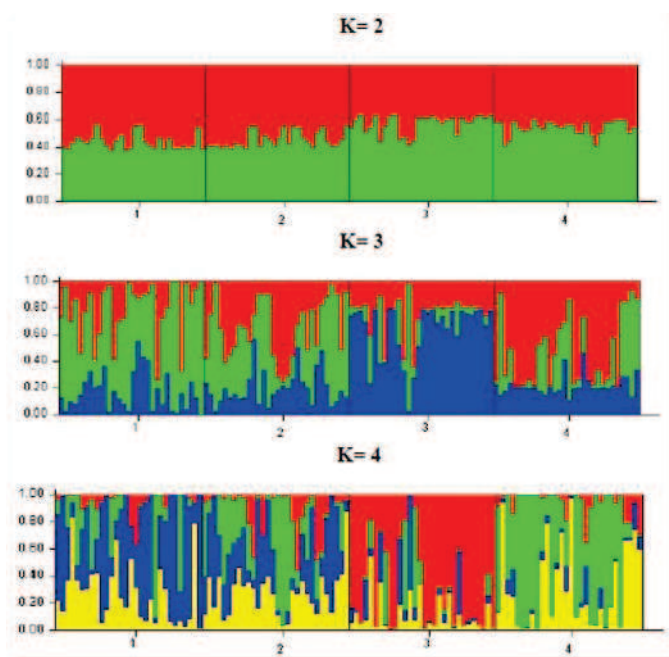


Figure IV.31. Structure génétique des populations de *Vitellaria paradoxa* établie selon le modèle « no admixture » du logiciel STRUCTURE (Pritchard et al. 2000) et en fonction du gradient Nord-Sud au Mali. K : nombre de groupes (clusters), 1 : Sassambourou, 2 : Daelan, 3 : Mperesso, 4 : Nafégué.

Deux variants sont donc observés à la position 331 de *VTE2-2* sur l'échantillon Mali : 331:A et 331:G. Sur l'ensemble des 76 séquences étudiées, 70 sont de type 331:A et seulement 6 sont du type 331:G. La distribution des deux variants alléliques est présentée à la **Figure IV.32**. On observe en particulier que les sites de Nafégué et Mperesso (les plus au sud du Mali), sont exempts du variant 331:G, et que celui-ci apparait uniquement au sein des deux sites les plus au Nord du gradient mais de façon minoritaire par rapport à l'allèle 331:A très abondant (**Figure IV.32**). On constate que le gain apporté par l'allèle 331:G sur la variation en α -tocophérol est fort. Ainsi les résultats obtenus lors du test d'association, estime un gain très négatif de -43.26 $\mu\text{g/g}$ d' α -tocophérol pour les haplotypes porteurs du variant allélique 331:A, par rapport au variant 331:G. Les boîtes à moustache présentée à la **Figure IV.33** confirment le gain apporté par le variant 331:G qui montre une moyenne en α -tocophérol très supérieur à celle enregistrée par l'autre variant allélique. Il convient cependant de garder à l'esprit que la position 331 du fragment de *VTE2-2* étudié se situe au sein d'un intron, le variant allélique 331:G ne constitue qu'une forme très minoritaire dans l'échantillonnage, que le pourcentage de variation expliqué n'est que très faible et que si l'association est significative au regard du test de Fisher au seuil 0.02, elle ne l'est plus après l'application d'une correction de Bonferroni (**Tableau IV.11**).

Chapitre IV : Étude du polymorphisme des gènes candidats

Tableau IV.11. Résultats des tests d'association réalisés avec le logiciel TASSEL (Bradbury *et al.*, 2007) entre les variants SNP du gène *VTE2-2* et les teneurs en α -tocophérol, β -tocophérol, γ -tocophérol et δ -tocophérol en utilisant 2 modèles d'analyse : avec et sans prise en compte de la structure génétique (Q). p = p-value du test d'exactitude de Fisher effectué sur le modèle ; R^2 = coefficient de détermination de la position du gène dans le modèle global tenant compte ou non de la structure ; * : valeurs significatives, p-value < 0,01

Tocophérol	Position	Modèle d'association structure $Y = \mu + L_{VTE2-2} + Q + \varepsilon$					Modèle d'association sans structure $Y = \mu + L_{VTE2-2} + \varepsilon$				
		p	R^2	Allèle 1	Gain de A1 sur A2	Allèle 2	p	R^2	Allèle 1	Gain de A1 sur A2	Allèle 2
α -tocophérol	37	0.193	0.014	C	33.55	A	0.185	0.024	C	42.69	A
	87	0.705	0.001	A	-3.57	G	0.328	0.013	A	11.16	G
	94	0.305	0.009	T	-16.58	A	0.828	0.001	T	4.12	A
	308	0.193	0.014	T	-33.55	G	0.185	0.024	T	-42.69	G
	331	0.015 *	0.041	A	-43.26	G	0.132	0.019	A	-31.45	G
	342	0.644	0.002	T	-12.60	G	0.947	0.000	T	2.15	G
	378	0.872	0.000	T	-1.45	G	0.193	0.023	T	14.09	G
	400	0.705	0.001	C	-3.57	T	0.328	0.013	C	11.16	T
	461	0.817	0.000	C	-2.10	T	0.218	0.020	C	13.41	T
	481	0.124	0.020	C	28.42	T	0.408	0.009	C	19.08	T
	559	0.553	0.003	C	10.92	A	0.419	0.009	C	18.62	A
	582	0.870	0.000	C	-4.32	G	0.659	0.003	C	14.31	G
	583	0.870	0.000	A	-4.32	T	0.659	0.003	A	14.31	T
	702	0.193	0.014	T	33.55	G	0.185	0.024	T	42.69	G
β -tocophérol	37	0.780	0.001	C	-0.34	A	0.957	0.000	C	0.08	A
	87	0.920	0.000	A	0.04	G	0.193	0.023	A	0.69	G
	94	0.642	0.002	T	0.35	A	0.321	0.013	T	0.88	A
	308	0.780	0.001	T	0.34	G	0.957	0.000	T	-0.08	G
	331	0.672	0.002	A	-0.37	G	0.985	0.000	A	-0.02	G
	342	0.636	0.002	T	-0.60	G	0.851	0.000	T	0.28	G
	378	0.971	0.000	T	-0.02	G	0.200	0.022	T	0.65	G
	400	0.920	0.000	C	0.04	T	0.193	0.023	C	0.69	T
	461	0.983	0.000	C	-0.01	T	0.199	0.022	C	0.65	T
	481	0.404	0.006	C	0.73	T	0.720	0.002	C	0.39	T
	559	0.600	0.002	C	0.45	A	0.458	0.007	C	0.80	A
	582	0.529	0.003	C	0.78	G	0.393	0.010	C	1.29	G
	583	0.529	0.003	A	0.78	T	0.393	0.010	A	1.29	T
	702	0.780	0.001	T	-0.34	G	0.957	0.000	T	0.08	G
γ -tocophérol	37	0.192	0.020	C	10.46	A	0.283	0.016	C	9.24	A
	87	0.511	0.005	A	1.92	G	0.768	0.001	A	0.90	G
	94	0.538	0.004	T	-3.10	A	0.083	0.040	T	-8.69	A
	308	0.192	0.020	T	-10.46	G	0.283	0.016	T	-9.24	G
	331	0.250	0.015	A	-6.61	G	0.320	0.013	A	-6.09	G
	342	0.510	0.005	T	-5.58	G	0.890	0.000	T	-1.20	G
	378	0.448	0.007	T	2.12	G	0.881	0.000	T	0.43	G
	400	0.511	0.005	C	1.92	T	0.768	0.001	C	0.90	T
	461	0.496	0.005	C	1.91	T	0.920	0.000	C	0.29	T
	481	0.602	0.003	C	3.02	T	0.425	0.009	C	4.89	T
	559	0.170	0.022	C	7.80	A	0.233	0.019	C	7.31	A
	582	0.956	0.000	C	-0.46	G	0.432	0.008	C	-6.77	G
	583	0.956	0.000	A	-0.46	T	0.432	0.008	A	-6.77	T
	702	0.192	0.020	T	10.46	G	0.283	0.016	T	9.24	G
δ -tocophérol	37	0.625	0.003	C	0.38	A	0.603	0.004	C	0.41	A
	87	0.674	0.002	A	0.12	G	0.303	0.014	A	0.29	G
	94	0.843	0.001	T	-0.10	A	0.901	0.000	T	-0.06	A
	308	0.625	0.003	T	-0.38	G	0.603	0.004	T	-0.41	G
	331	0.946	0.000	A	-0.04	G	0.838	0.001	A	0.12	G
	342	0.994	0.000	T	-0.01	G	0.603	0.004	T	0.41	G
	378	0.549	0.005	T	0.16	G	0.238	0.019	T	0.32	G
	400	0.674	0.002	C	0.12	T	0.303	0.014	C	0.29	T
	461	0.565	0.004	C	0.16	T	0.246	0.018	C	0.31	T
	481	0.442	0.008	C	0.43	T	0.458	0.007	C	0.42	T
	559	0.562	0.004	C	0.32	A	0.458	0.007	C	0.42	A
	582	0.598	0.004	C	0.42	G	0.603	0.004	C	0.41	G
	583	0.598	0.004	A	0.42	T	0.603	0.004	A	0.41	T
	702	0.625	0.003	T	0.38	G	0.603	0.004	T	0.41	G

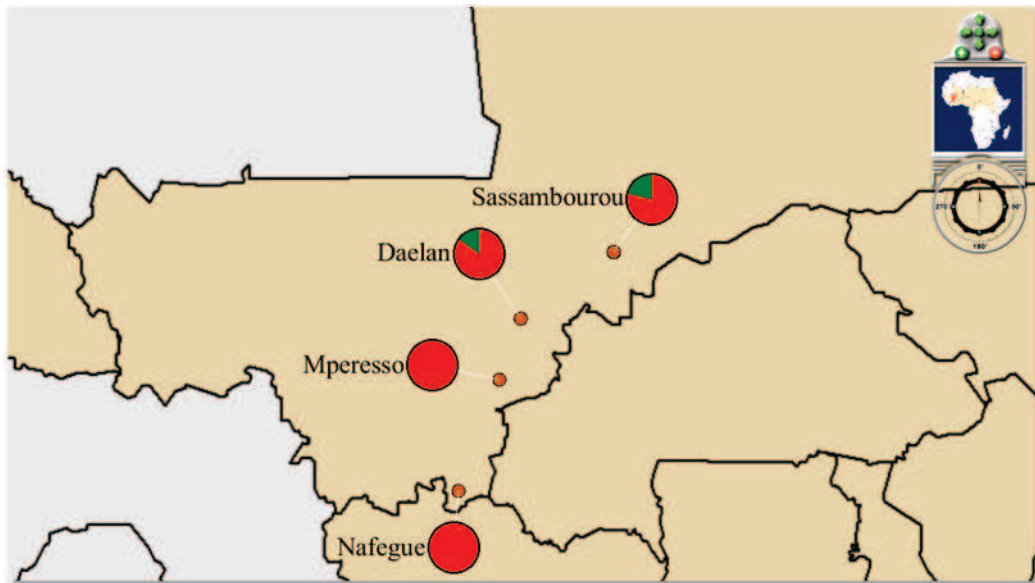


Figure IV.32. Distribution des allèles 331:A (en rouge) et 331:G (en vert) au sein de l'échantillon Nord-Sud au Mali

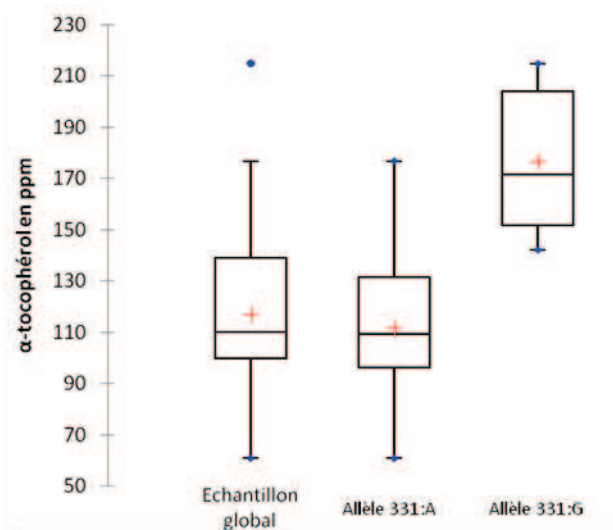


Figure IV.33. Boîtes à moustache synthétisant la variation des valeurs d' α -tocophérol : au sein de l'échantillon global ; au sein des individus porteurs de l'allèle *VTE2-2_331:A* ; et au sein des individus porteurs de l'allèle *VTE2-2_331:G*. Les croix rouges représentent la moyenne et les points bleus les valeurs extrêmes.

4 DISCUSSION

4.1 Le niveau de ploïdie de *Vitellaria paradoxa*

Faisant état de la duplication du gène *SAD* chez *Vitellaria paradoxa*, nous avons entrepris une étude cytogénétique rapide de quelques accessions de types *paradoxa* et *nilotica*. Si nous avons pu vérifier que les deux sous-espèces sont bien diploïdes, nous avons constaté 13 paires de chromosomes (**Figure IV.7**). Or, une étude de Miège en 1954 avait comptabilisé seulement 12 paires de chromosomes sur des spécimens d'origine de Niangoloko au Burkina Faso (Miège, 1954; Hall *et al.*, 1996). Cette observation a ensuite été confirmée par Pilz en 1980 sur des échantillons en provenance de Kwara au Nigéria (Pilz, 1980) et par Lovett (résultats non publié). Miège cependant reconnaît que la petite taille et l'épaisseur des chromosomes de Karité rendent leur comptage relativement difficile (Hall *et al.*, 1996). Les résultats que nous avons obtenus sont sur les accessions des deux sous-espèces sont donc clairement en contradiction avec la littérature. Il conviendra donc de pousser plus avant les investigations du nombre et de la taille des chromosomes des deux sous-espèces de Karité.

4.2 Une duplication récente du gène *SAD* chez *Vitellaria paradoxa*

Lors de ce travail de thèse, nous nous sommes en particulier intéressés au patron de variabilité du gène *SAD*. Cependant, le clonage puis le séquençage du gène chez *Vitellaria paradoxa* a permis de mettre en évidence la présence de deux copies extrêmement proches du gène, homologues à 99%, que nous avons désigné sous les noms de *SAD1* et *SAD2*. Les événements de duplication de gènes ont été beaucoup étudiés et commentés dans la sphère des biologistes de l'évolution des espèces permettant d'établir un certain nombre de théories.

Ainsi, les premières observations de duplication de gènes ont été rapportées dès 1936 avec les travaux de Bridges, constatant le doublement d'une bande d'un des chromosomes d'un mutant de *Drosophila melanogaster* (Bridges, 1936). Ce mécanisme apparaît jusqu'avant les années 90, comme un événement rare (Ohno, 1967; Nei, 1969; Ohno, 1970). Cependant, avec l'avènement de la biologie moléculaire et les séquençages de génomes complets, la communauté scientifique réalise alors l'importante prévalence des



Chapitre IV : Étude du polymorphisme des gènes candidats

gènes dupliqués au sein de tout organisme. Les duplications de gènes apparaissent comme un des principaux moteurs de l'évolution des génomes, donc des espèces (Friedman & Hughes, 2001; Wolfe, 2001; Zhang *et al.*, 2001; Spring, 2002; Zhang, 2003), La duplication de gènes constituant un tremplin pour l'acquisition de nouvelles fonctions et la génération de nouvelles adaptations (Force *et al.*, 1999; Lynch & Force, 2000; Moore & Purugganan, 2005). En ce sens, les évènements de duplication contribuent significativement à l'augmentation de la complexité du fonctionnement des organismes (Freeling & Thomas, 2006).

Selon le modèle classique de Ohno (1970), après la survenue d'une duplication de gène, les contraintes sélectives se concentrent sur une copie. La seconde copie peut alors accumuler des mutations, et la conséquence la plus courante est la perte de fonction de cet exemplaire du gène. Dans de rares cas, cependant, une mutation peut être avantageuse dans un environnement spécifique et conduire alors à une néo-fonctionnalisation du gène (Beisswanger & Stephan, 2008). Afin d'expliquer le maintien de gènes dupliqués fonctionnels, Hughes propose une théorie selon laquelle, si le gène ancestral (avant duplication) assurait une ou plusieurs fonctions, les copies naissantes entreprennent alors de se spécialiser vers des sous-fonctions complémentaires, appelées sous-fonctionnalisations (Hughes, 1994).

Chez les plantes, une part importante de la diversité génétique résulte de la duplication des gènes suivie de spécialisation adaptative des deux copies (Flagel & Wendel, 2009). Le séquençage du génome du peuplier en 2006 (Tuskan *et al.*, 2006) permet d'identifier plus de 45000 gènes et parmi eux, environ 8000 paires de gènes dupliqués chez *Populus trichocarpa*. Ainsi, pour chaque gène identifié chez *Arabidopsis thaliana*, une moyenne de 1.5 gènes homologues fonctionnels sont mis en évidence chez le peuplier (Tuskan *et al.*, 2006). Ceci s'explique par un évènement de duplication totale du génome du peuplier suivi d'un réarrangement chromosomique majeur et d'une re-diploïdisation du génome de *Populus trichocarpa*. L'émergence de ces copies de gène a en outre eu un impact très important dans la dynamique évolutive du peuplier. Ainsi, cet évènement aura permis le renforcement de processus fonctionnels majeurs pour les ligneux, avec le maintien de 18 gènes au sein de la famille multigénique *CesA* (contre seulement 10 chez *Arabidopsis thaliana*) indispensables à la production des parois lignocellulosiques. De plus, cet évènement est à l'origine de l'apparition de nombreux nouveaux gènes et de nouvelles



Chapitre IV : Étude du polymorphisme des gènes candidats

fonctions métaboliques, notamment dans les voies de signalisation moléculaire et les mécanismes de résistance aux maladies (Tuskan *et al.*, 2006). Les événements de duplication sont ainsi à l'origine de nombreuses familles multigéniques chez les plantes. C'est le cas notamment de la famille des acyl-acyl-désaturases à laquelle le gène *SAD* appartient (Kachroo *et al.*, 2007). Cependant, la duplication observée du gène *SAD* chez *Vitellaria paradoxa* reste à notre connaissance inédite. Quel impact la duplication du gène *SAD* chez *Vitellaria paradoxa* a-t-elle sur la biologie de l'espèce ? Quelle dynamique évolutive est-elle associée à ces deux copies *SAD1* et *SAD2* ?

Au sein de ce travail de thèse, nous mettons en évidence que la duplication du gène *SAD* révèle deux patrons distincts en fonction de la sous-espèce considérée.

D'une part chez la sous-espèce *paradoxa*, la copie *SAD2* est non fonctionnelle (**Figure IV.8**). Ainsi, bien que le gène *SAD2* soit transcrit chez la sous-espèce, la traduction de l'ARNm résulte en une protéine tronquée. Cette copie non-fonctionnelle montre des niveaux de polymorphisme plus importants que ceux observés sur la copie *SAD1* chez la sous-espèce *paradoxa*. Ainsi, renforcé par le constat d'une homogénéisation des valeurs de π entre les introns et les exons, *SAD2* apparaît donc comme une copie en cours de divergence pour la sous-espèce. Cette accumulation de mutation au sein de *SAD2* associée au relâchement des contraintes sélectives, est cohérent avec le modèle théorique proposé par Ohno (Ohno, 1970). Si la divergence entre *SAD1* et *SAD2* se confirme chez la sous-espèce *paradoxa*, on peut donc s'attendre à une évolution de *SAD2* vers un pseudogène (Vanin, 1985; Balakirev & Ayala, 2003). Il est cependant possible que cette séquence brut soit à l'origine d'un nouveau gène au sein de la famille multigénique des acyl-acyl-désaturases, avec peut-être l'émergence d'une nouvelle fonction ou d'une délocalisation cellulaire (Beisswanger & Stephan, 2008).

D'autre part, nous avons vu que les deux copies transcrites du gène *SAD* donnent *in silico* des protéines fonctionnelles *SAD1* et *SAD2* au sein de la sous-espèce *nilotica*. Ainsi, contrairement à ce que nous venons de voir pour la sous-espèce *paradoxa*, la dynamique évolutive du gène *SAD2* chez *nilotica* est sensiblement similaire à celle de la copie *SAD1*, avec un polymorphisme faible. Le maintien de ces deux copies sous une forme fonctionnelle pourrait ainsi apporter un bénéfice physiologique et constituer une stratégie adaptative



pour *V. p. nilotica*. En effet, nous avons fait état, en début de ce chapitre, du rôle adaptatif important de la protéine SAD, dont la fonction est d'induire la désaturation de l'acide stéarique en acide oléique (Somerville *et al.*, 2000). De plus, la surexpression du gène chez *Arabidopsis thaliana* induit, en particulier, des niveaux d'acide oléique plus importants (Kachroo *et al.*, 2007) et on peut supposer qu'il en est de même chez le Karité. D'autre part nous avons vu dans le Chapitre III de cette thèse, que la sous-espèce *nilotica* montre des proportions d'acide oléique très supérieures à celles observées chez la sous-espèce *paradoxa*. Il est donc possible que la présence de deux copies fonctionnelles chez *nilotica* soit responsable de cette divergence phénotypique. Enfin, le maintien de deux copies fonctionnelles chez la sous-espèce *nilotica*, évoluant dans un environnement sensiblement plus froid et plus humide que la sous-espèce *paradoxa*, pourrait constituer un avantage adaptatif dans cette région (Linder, 2000).

Le gène *SAD2* pourrait ainsi constituer le marqueur d'une divergence adaptative voire d'une spéciation future entre les sous-espèces *paradoxa* et *nilotica*.

4.3 Une sélection directionnelle des gènes étudiés chez *Vitellaria paradoxa* à l'échelle de l'aire naturelle

Au cours de ce Chapitre, l'analyse des indices de diversité, des réseaux haplotypiques et des arbres phylogénétiques met en évidence un très faible niveau de polymorphisme au sein des séquences étudiées et une divergence très faible entre les populations. L'étude des patrons de sélection indique que cette diversité faible pourrait résulter d'une sélection directionnelle aux loci étudiés à l'échelle de l'aire naturelle. Cependant, à l'échelle de l'aire de distribution du Karité, nous avons montré au cours du Chapitre II de cette thèse, qu'il existe une forte structuration des populations et une différenciation significative entre les trois grands groupes que constituent l'Afrique de l'Ouest, l'Afrique Centrale et l'Afrique de l'Est. Ainsi, il est probable que la mise en évidence d'effet de sélection au sein de l'échantillon global soit fortement impactée par des effets de structure génétique. Dans la littérature, Krutovsky et Neale mettent en évidence les traces de sélection positive au sein de l'espèce *Pseudotsuga menziesii* (le Douglas), mais attribuent ces résultats négatifs de Tajima à la conséquence d'une expansion récente des populations (Krutovsky & Neale, 2005). L'étude de la diversité de 5 gènes au sein de peuplements de *Populus tremula*



Chapitre IV : Étude du polymorphisme des gènes candidats

d'Europe du Nord et d'Europe Centrale montre que les processus démographiques semblent, plus que la sélection, la cause de l'excès de polymorphismes de faibles fréquences (Ingvarsson, 2005).

Au sein de notre étude, nous mettons en évidence que les traces de sélection observées à l'échelle de l'aire naturelle ont tendance à s'effacer au niveau des sous-groupes échantillonnés. En particulier c'est systématiquement le cas pour les échantillons d'Afrique de l'Est et d'Afrique Centrale, pour lesquels nous n'avons pas détecté de sélection significative au sein des différents fragments étudiés (**Tableaux IV.4 et IV.6**). Il est donc difficile pour ces deux groupes d'émettre des hypothèses sur la sélection des gènes étudiés. Il est cependant également probable que ce résultat soit lié à un manque de puissance inhérent au faible nombre de sites polymorphes et de séquences obtenues au sein des régions Est et Centre.

Nous mettons en évidence des excès de variants rares au sein des différentes portions de gène pour l'échantillon d'Afrique de l'Ouest. Au cours de cette thèse nous avons évoqué deux hypothèses pouvant expliquer ce type de résultat : l'impact d'une sélection positive au sein de la population ou l'effet collatéral d'un processus démographique affectant cette population. Nous discutons donc les deux hypothèses en lien avec les données obtenues sur les individus d'Afrique de l'Ouest, dans les paragraphes suivants.

Ainsi, concernant le gène *SAD1*, nous avons observé en Afrique de l'Ouest des valeurs significatives et négatives pour les tests de Tajima ($D_{SAD1} = -2$; $p < 0.02$) (Tajima, 1989) et de Fu et Li ($D^*_{SAD1} = -3$; $F^*_{SAD1} = -3.2$; $p < 0.02$) (Fu & Li, 1993) (**Tableau IV.4 et Figure IV.12**). Ces résultats indiquent donc un excès de variant polymorphes rares, compatibles avec les effets d'une sélection directionnelle au voisinage des introns du gène *SAD1*. Sous une telle sélection, le balayage sélectif entraîne une diminution drastique de la diversité expliquant ainsi le faible polymorphisme observé. Les résultats obtenus laissent donc supposer que ce gène est soumis à une pression de sélection en Afrique de l'Ouest. Sous cette hypothèse, il est possible que cette sélection soit liée au rôle adaptatif important du contrôle des proportions relatives d'acides gras (Linder, 2000). Il est également possible que cette sélection résulte de la domestication de l'espèce qui remonterait à près de 1000 ans (Lovett & Haq, 2000; Maranz & Wiesman, 2003), et se serait exercée principalement en Afrique de



Chapitre IV : Étude du polymorphisme des gènes candidats

l'Ouest et en particulier sur le plateau de Mossi au Burkina Faso (Lovett & Haq, 2000; Maranz & Wiesman, 2003).

Les résultats obtenus pour *SAD2*, copie non fonctionnelle du gène *SAD*, suggèrent également des traces de sélection directionnelle pour l'échantillon d'Afrique de l'Ouest. Ce résultat apparaît en contradiction avec l'hypothèse d'une dérive du gène *SAD2*. Cependant, nous pouvons penser que cette copie divergente ait un rôle important dans un processus de « gene silencing » (extinction de gène) par petits ARN interférents (si RNA – « small interfering RNA »). Ce mécanisme apparaît comme une alternative à la dérive d'une copie de gène consécutive à un évènement de duplication (Rodin & Riggs, 2003). Ce type de processus, récemment mis en évidence, permet une régulation épigénétique de la production d'une protéine par contrôle post-transcriptionnel (empêchant la traduction de l'ARNm en protéine) (Hamilton & Baulcombe, 1999; Gura, 2000). Ainsi, sous ce scénario, il serait possible que la copie *SAD2* ait un rôle dans le contrôle épigénétique de l'expression de *SAD1*. En outre cette hypothèse n'est pas impossible car nous avons constaté la transcription de *SAD2* au sein de la sous-espèce *paradoxa*. Une fois transcrite, l'ARNm issue de la copie *SAD2*, n'a d'autre destinée que celle d'être dégradée car la traduction protéique est impossible chez cette copie tronquée (**Figure IV.9**). Une fois clivé en petit fragment sous l'action d'une endonucléase de type III, appelée « Dicer », les petits fragments de cet ARNm (les RNAsi de 20 à 25 pb), peuvent ainsi être pris en charge par un complexe multiprotéique appelé RISC (RNA-Induced Silencing Complex), et servir comme séquence de reconnaissance pour le clivage d'ARNm de la copie *SAD1* présents dans le cytoplasme (Meister & Tuschl, 2004). Ceci entraînerait alors une désaturation moins efficace des acides gras saturés pour les individus porteurs de la copie *SAD2* tronquée. Cette hypothèse est donc cohérente avec les patrons de composition chimique des graines de karité observés, fortement saturés au sein de la sous-espèce *paradoxa* (copie *SAD2* tronquée) et fortement désaturés au sein de la sous-espèce *nilotica* (copie *SAD2* fonctionnelle). Il devient alors possible que la copie *SAD2* ait un impact fort sur l'expression de la composition des graines, et pourrait ainsi être soumise à la sélection, ce qui expliquerait le patron de sélection observé.

Enfin, le gène *VTE2-2*, à l'origine de la traduction de la protéine homogénisate phytyltransférase, enzyme clef de la voie de biosynthèse de tocophérols indique également un excès de variants polymorphes rares, suggérant une sélection directionnelle du gène à



l'Ouest de l'aire naturelle (**Tableau IV.6** et **Figure IV.24**). Ainsi, de même que pour *SAD1* et *SAD2*, la sélection positive serait à l'origine d'une réduction drastique de la diversité de *VTE2-2* et suppose que le fragment étudié est à proximité d'une région fonctionnellement importante.

Les patrons de sélection mis en évidence au sein de l'Afrique de l'Ouest pour les différents gènes, convergent donc vers l'hypothèse d'une sélection directionnelle. Or selon les modèles établis, la sélection ne s'exprime pas de façon homogène au sein des génomes, mais cible que quelques loci en particulier impliqués dans la valeur sélective (Cavalli-Sforza, 1966). On s'attend donc à ce qu'une majorité de loci ne soit pas soumise à sélection. Il est donc possible, de même que ce que nous avons suggéré à l'échelle de l'aire naturelle, que les résultats obtenus à l'échelle de l'Afrique de l'Ouest soient liés à des contraintes démographiques. Nous avons en outre évoqué dans ce Chapitre que l'histoire démographique des populations peut aller jusqu'à mimer certains effets de la sélection (Schlotterer, 2003; Stinchcombe & Hoekstra, 2008) (**Figure IV.1**). Nous avons également mis en évidence par des simulations de coalescence dans le Chapitre II de cette thèse (**Box II.1**), que les populations de Karité en Afrique de l'Ouest sont en expansion démographique. Or, une population en expansion démographique pourra mimer les effets d'une sélection directionnelle (Stinchcombe & Hoekstra, 2008) (**Figure IV.1**). Ainsi, comme montré chez d'autres espèces d'arbres (Ingvarsson, 2005; Krutovsky & Neale, 2005), l'expansion des populations en Afrique de l'Ouest pourrait donc être à l'origine des patrons de sélections évoqués par nos gènes.

4.4 Une adaptation locale de *Vitellaria paradoxa* au Nord du Mali

L'étude des patrons de variabilité nucléotidique au sein du gène *VTE2-2* pour les sites du gradient Nord-Sud du Mali met en évidence des traces significatives de sélection directionnelle (**Tableau IV.8**). Cette sélection n'est cependant plus significative au sein des sites de Nafégué, Mperesso et Daelan (**Tableau IV.8** et **Figure IV.28**). En revanche le signal détecté est fort pour le site de Sassambourou à l'extrême nord du Mali. Cette observation est donc compatible avec l'impact d'une sélection positive en ce site. L'analyse de la structuration des sites étudiés par les marqueurs neutres microsatellites indique une faible structuration génétique, et donc des flux de gènes importants. Ainsi, le patron de sélection



Chapitre IV : Étude du polymorphisme des gènes candidats

observé rend compte d'une adaptation locale des individus de Sassambourou. Cette adaptation pourrait ainsi résulter du climat extrême qui caractérise Sassambourou à la limite nord de l'aire naturelle avec des températures élevées et des niveaux faibles de pluviométrie. La description des sites du gradient Mali est présentée dans le Chapitre III de cette thèse.

La mise en œuvre d'étude d'association est dépendante du niveau de déséquilibre de liaison (DL) observé au sein du génome de l'espèce étudiée. Ainsi, plus le DL est élevé, plus la taille de la région caractérisée par un marqueur ciblé sera importante, couvrant parfois plusieurs gènes. En revanche, il est alors difficile (voire impossible) de localiser précisément le polymorphisme responsable de la variation phénotypique observée. Ainsi, c'est souvent le cas pour les espèces de grandes cultures autogames et ayant une histoire ancienne de domestication comme notamment le *Sorghum bicolor* (Sorgho) (Hamblin *et al.*, 2004). Contrairement à ces espèces de grandes cultures, on observe des valeurs faibles de DL au sein des espèces forestières. Ces faibles valeurs de déséquilibre de liaison sont notamment dues au régime essentiellement allogame de ces espèces, au nombre important de générations qu'elles comptent et aux recombinaisons fréquentes au sein de leurs génomes. Si un marqueur donné ne permettra pas de caractériser une grande région du génome, le marquage répété de segments chromosomiques de petite taille permet de repérer plus facilement le polymorphisme associé avec la variation phénotypique (Neale & Savolainen, 2004).

Aucune mesure de déséquilibre de liaison n'a à ce jour été rapporté pour l'espèce *Vitellaria paradoxa*. Ainsi, les résultats de DL mesurés au cours de cette thèse constituent donc une avancée importante de ce point de vue. En outre nous avons montré que le déséquilibre de liaison est très faible au sein de l'aire naturelle, pour l'ensemble des gènes étudiés (**Figures IV.10, IV.11 et IV.23**). En revanche, à une échelle locale au Mali, nous avons mis en évidence que le DL n'est pas forcément homogène au sein d'une population. Nous avons montré un DL particulièrement fort au sein du gène *VTE2-2*, pour le site de Sassambourou ($R^2 = 0.67$) par rapport à celui observé au sein de Daelan ($R^2 = 0.17$) très proche géographiquement (**Figure IV.27**). Ce déséquilibre de liaison fort à Sassambourou constitue en outre un indice supplémentaire renforçant l'hypothèse d'une adaptation locale des individus à ce site.



Chapitre IV : Étude du polymorphisme des gènes candidats

Les tests d'association effectués avec le modèle tenant compte de la structure des individus du gradient Nord-Sud au Mali, nous ont permis de mettre en évidence une association significative ($p = 0.015$) entre les teneurs en α -tocophérol et la position 331 sur le gène de *VTE2-2*. Cette association révèle l'existence de deux variants alléliques. Le premier variant (331:A), majoritaire au sein de l'échantillon apparaît favorable à des teneurs plus faibles en α -tocopérol, tandis que le second (331:G) semblerait associer aux phénotypes de fortes teneurs en α -tocophérol (**Tableau IV.11** et **Figure IV.33**). La position 331 se situe dans l'intron 3 du gène *VTE2-2*, au sein de « haplobloc » mis en évidence pour les échantillons de Daelan et Sassambourou (**Figure IV.27**). L'impact du polymorphisme intronique sur la variabilité phénotypique chez les plantes semble faible, et a été surtout abordée dans le cas des espèces cultivées (Deyholos & Sieburth, 2000; Fiume *et al.*, 2004). Il existe en outre peu d'information sur le polymorphisme des zones introniques dans le génome des espèces forestières. Ainsi, certaines études proposent que les introns puissent jouer un rôle important dans les régulations de transcription de gène. En particulier, l'étude du gène *VRN-1*, impliqué dans la réponse à la vernalisation des plantes, montre qu'une délétion au sein du premier intron du gène entraîne une variation importante dans le niveau de transcription du gène chez *Triticum aestivum* L. (le blé) (Rhone *et al.*, 2008). Cependant, il est probable que le polymorphisme observé à la position 331 ne soit que l'indicateur d'un polymorphisme fonctionnel proche. Ainsi, compte tenu du faible DL global mis en évidence, nous pouvons envisager une étude plus approfondie au voisinage du fragment étudié. Plus encore, on observe que le variant bénéfique aux teneurs fortes en α -tocophérols se situe uniquement au sein des deux populations les plus au nord du gradient échantillonné, et en particulier à Sassambourou. Ce résultat constitue donc un argument supplémentaire à l'hypothèse d'une adaptation locale. Cependant, l'association détectée est à relativisée tenant compte du relativement faible échantillon utilisé. Également, si la significativité de l'association est bonne au regard du test de Fisher utilisé par défaut par la procédure GLM du logiciel TASSEL (Yu *et al.*, 2006; Bradbury *et al.*, 2007), l'association mise en évidence n'est plus significative après l'application d'une correction statistique de Bonferroni.

5 CONCLUSION ET PERSPECTIVES

Au cours de ce chapitre nous avons mis en évidence la duplication du gène *SAD* chez *Vitellaria paradoxa* constituant une nouvelle observation au sein de la famille multigénique des acyl-acyl-désaturases. Nous notons que les dynamiques évolutives des gènes *SAD1* et *SAD2* ne sont pas homogène au sein de l'espèce. Ainsi dans notre échantillonnage, au sein de la sous-espèce *paradoxa*, la traduction *in silico* du gène *SAD2* suggère la production d'une protéine tronquée donc non fonctionnelle. Dans ce contexte, nous suggérons qu'une accumulation progressive de mutation au sein de la séquence pourrait *in fine* mener un pseudogène ou éventuellement à l'acquisition d'une nouvelle fonction à partir de ce matériel génétique brut. À l'inverse, on montre que les deux copies de gène sont potentiellement fonctionnelles au sein de la sous-espèce *nilotica*. Nous proposons ainsi que ce doublement du gène pourrait être à l'origine des fortes teneurs en acide oléique au sein de la sous-espèce (cf Chapitre III). Bien que la distinction des sous-espèces ne soit pas claire à l'heure actuelle (Chapitre II), il est possible que ce gène *SAD2* soit le reflet d'un processus de spéciation en cours entre les deux sous-espèces

Afin de tester ces hypothèses, nous envisagerons deux approches :

La première approche correspond à une démarche de génomique fonctionnelle. Ainsi, l'étude du niveau d'expression des copies *SAD1* et *SAD2*, et du gène *VTE2-2* en milieu contrôlé chez des individus de type *paradoxa* et *nilotica*, nous permettrait de porter un regard quantitatif sur les variations observées sur le plan de l'architecture simple des séquences. La transformation de lignées mutantes *sad (ssi2)* d'*Arabidopsis thaliana* (Kachroo *et al.*, 2003b) avec les différentes versions « *nilotica* » et « *paradoxa* » du gène *SAD2*, nous permettrait d'une part de vérifier la fonctionnalité protéique de *SAD2*, et d'autre part de mieux comprendre l'impact éventuel du doublement du gène *SAD* sur le phénotype des lignées transformantes.

Dans une approche évolutionniste, nous proposons de tester l'hypothèse que *SAD2* constituerai le marqueur d'une potentielle spéciation entre *paradoxa* et *nilotica*. Pour cela, il conviendrait d'une part de séquencer le gène *SAD2* sur un panel plus important d'individus mieux répartis sur l'aire naturelle, en incluant notamment des individus d'Afrique centrale et

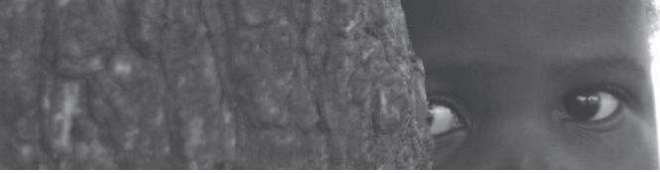


Chapitre IV : Étude du polymorphisme des gènes candidats

plus d'individus d'Afrique de l'Est. D'autre part il serait intéressant de comparer les patrons de diversités de *SAD2* avec ceux de gènes candidats potentiellement responsables de l'isolation reproductive de deux populations ou deux sous-espèces. Ce champ exploratoire récent de recherche de gènes dits « de spéciation » a permis d'identifier de nombreux loci candidats chez les plantes (Rieseberg & Blackman, 2010) tels que *AN2* (Quattrocchio *et al.*, 1999; Hoballah *et al.*, 2007), *HPA1* (Bikard *et al.*, 2009) ou encore *HBD2* (Yamamoto *et al.*, 2010), qui pourront ainsi être étudiés.

Nous mettons en évidence dans ce Chapitre les traces d'une possible adaptation locale au Nord Mali pour le gène *VTE2-2*, qui pourrait être à l'origine de teneurs accrues en tocophérols apportant une valeur adaptative aux arbres évoluant en milieu sec et chaud. L'analyse de l'association entre les variants alléliques du gène *VTE2-2* et les teneurs en tocophérols au sein du gradient Nord-Sud au Mali, nous permet de mettre en évidence une association significative entre les teneurs en α -tocophérols et le site 331 du gène *VTE2-2*, qui reste cependant à être vérifiée.

Pour ce type de test, on préconise, notamment pour les espèces cultivées dont on connaît généralement bien la structure génétique et l'histoire évolutive, l'utilisation de 250 à 400 lignées, et de plusieurs gènes candidats (Camus-Kulandaivelu *et al.*, 2006; Zhao *et al.*, 2007; Cockram *et al.*, 2008). Afin de confirmer ce résultat, nous proposons d'analyser le génotype d'un plus grand nombre d'individus du gradient, d'amplifier la longueur du fragment considéré et de multiplier le nombre de gènes (Oraguzie *et al.*, 2007; Ingvarsson *et al.*, 2008; Eckert *et al.*, 2009).



Discussion Générale

L'objectif scientifique général de cette thèse était de mieux comprendre l'impact combiné des facteurs évolutifs sur la variabilité au sein de *Vitellaria paradoxa*. Afin de répondre au mieux à cet objectif nous avons effectué une étude de phylogéographie combinant les informations de différents types de marqueurs (microsatellites nucléaires, microsatellites chloroplastiques, séquences intergéniques du chloroplaste) caractérisés par des taux de mutations différents (Wolfe *et al.*, 1987; Vigouroux *et al.*, 2002). Pour mener à bien cette étude, nous nous sommes basés sur un échantillonnage large, réparti sur l'aire naturelle de *Vitellaria paradoxa* (71 populations ; 364 arbres génotypés) et avons considéré les caractéristiques climatiques de cette région à travers diverses périodes de l'ère quaternaire (moderne ; dernier maximum glaciaire ; et de la dernière période interglaciaire). Nous nous sommes ensuite intéressés aux patrons de variation chimique des graines de karité (acides gras et tocophérols) à travers un échantillonnage de 624 arbres répartis sur 17 sites. Nous avons choisi deux axes d'étude : différenciation chimique entre les deux sous-espèces *nilotica* et *paradoxa* ; et variabilité chimique des graines le long de gradients climatiques. Enfin, nous nous sommes focalisés sur l'étude du polymorphisme de gènes (*SAD* et *VTE2-2*) impliqués dans les voies de biosynthèse des constituants chimiques étudiés. Dans cette approche, nous avons mis en place une stratégie d'étude de la diversité nucléotidique, d'analyse du déséquilibre de liaison, de recherche de signature de sélection et de tests d'association génotype/phénotype s'intéressant d'une part à un échantillon représentatif des deux sous-espèces et d'autre part à un échantillon réparti le long d'un gradient climatique Nord-Sud au Mali. Dans ce dispositif, en milieu environnemental non contrôlé, nous proposons une réflexion tenant compte de différents effets confondus car les contraintes expérimentales ne nous ont pas permis de dissocier les effets génétiques, les effets environnementaux et les effets d'interaction génotype X environnement.

La stratégie méthodologique suivie, combinant l'utilisation de marqueurs neutres, de gènes et des composantes de caractères complexes, nous a permis d'avoir une vision globale des patrons de variation de *Vitellaria paradoxa*. Avant de discuter de ces patrons, nous rappelons tout d'abord les principaux résultats mis en évidence en relation avec les hypothèses de départ formulées dans cette thèse.

1 RAPPEL DES PRINCIPAUX RÉSULTATS ET DES HYPOTHÈSES TESTÉES

1.1 Histoire phylogéographique de *Vitellaria paradoxa*

Rappel de l'hypothèse testée (1) :

Les changements climatiques liés aux dernières grandes glaciations ont eu un impact marqué sur la distribution de la diversité génétique.

Au cours de cette étude, nous avons montré que la diversité génétique du Karité est relativement faible pour les séquences intergéniques du chloroplaste, et que ces marqueurs ne sont pas bien adaptés pour décrire les patrons de variation de l'espèce même sur son aire naturelle. Cependant, nous avons noté que cette faible variabilité n'est pas exceptionnelle au sein d'espèces d'arbres à grande aire de répartition (Vendramin *et al.*, 2008; Pock Tsy *et al.*, 2009).

Nous avons mis en évidence que la structure génétique de l'espèce suit un profil phylogéographique, que ce soit avec les marqueurs microsatellites chloroplastiques ($G_{ST} = 0.81$; $N_{ST} = 0.90$; $N_{ST} > pN_{ST}$; **Figures II.3 et II.4**), ou nucléaires ($F_{ST} = 0.21$; $R_{ST} = 0.28$; $R_{ST} > pR_{ST}$; **Figures II.5 et II.6**). En utilisant les résultats de modélisation des niches écologiques (ENM) de l'espèce (Hijmans *et al.*, 2001), nous proposons que, si l'impact de l'homme à travers une potentielle domestication de l'espèce il y a 1000 ans en Afrique de l'Ouest (Neumann *et al.*, 1998; Lovett & Haq, 2000a; Maranz & Wiesman, 2003) peut expliquer en partie les patrons de diversité génétique observés (et notamment le phénomène d'expansion en Afrique de l'Ouest), ceux-ci sont surtout cohérents avec un fort impact des perturbations climatiques au dernier maximum glaciaire (**HYPOTHÈSE 1**) (**Figure II.7**). En outre, nous montrons que ces perturbations du Quaternaire ne sont pas intervenues de façon homogène sur l'aire naturelle de l'espèce, mais ont surtout affecté l'Afrique de l'Est où elles auraient entraîné une fragmentation importante et une réduction drastique des peuplements, nous permettant de proposer des refuges potentiels où l'espèce aurait été restreinte (**Figure II.7**). Nos prédictions suggèrent par ailleurs que les populations d'Afrique de l'Ouest auraient été relativement épargnées par ces changements climatiques, expliquant ainsi la présence d'un « hotspot » de diversité dans cette région (**Figure II.2**). Nous proposons que les patrons de différenciation liés à ces événements climatiques auraient été

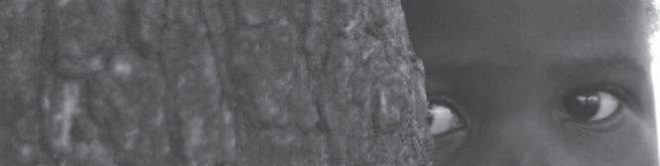
maintenus par des effets d'isolement par la distance et par la présence du massif Adamawa à la frontière entre l'Afrique Centrale et l'Afrique de l'Ouest. Nous proposons dans cette étude que la différenciation génétique observée à l'échelle de l'aire naturelle est en contradiction avec la différenciation taxonomique entre les deux sous-espèces (**Figures II.3, II.4, II.5 et II.6**). Nous suggérons enfin une possible origine des populations actuelles de *Vitellaria paradoxa* en Afrique de l'Ouest, cohérente avec les résultats obtenus pour le baobab (*Adansonia digitata*), une autre espèce d'arbre de Savane avec une distribution similaire au Karité en Afrique sub-sahélienne (Pock Tsy *et al.*, 2009).

1.2 Variabilité chimique des graines de Karité

Rappel de l'hypothèse testée (2) :

Les variations de constituants chimiques des graines de Karité (acides gras et tocophérols) suivent un cline environnemental (lié principalement à la température et à la pluviométrie).

Dans cette étude nous avons confirmé les résultats proposés par Maranz *et al.* (2004) en observant de très fortes teneurs en matière grasse dans les graines de Karité. Cette matière grasse représente ainsi entre 30 et 60% de la matière sèche totale (**Figure III.8**) ce qui est comparable aux fèves de cacao (50%) (*Theobroma cacao* L.) (Liendo *et al.*, 1997). Les résultats obtenus sont également comparables à ceux de Maranz *et al.* (2004) pour les proportions relatives des acides gras (**Figure III.12**) mais diffèrent sensiblement pour les teneurs en tocophérols (**Figure III.13**). Nous proposons en outre que les teneurs en tocophérols ont peut-être été un peu surestimées dans les études de Maranz et Wiesman (2004), car dans notre cas nous avons utilisé des répétitions pour le dosage de ces composés présents en faibles quantités. Cependant, conformément aux résultats antérieurs, nous observons une différenciation forte entre les deux sous-espèces pour ces constituants. En particulier, nous montrons que les proportions relatives d'acide oléique sont plus fortes au sein de la sous-espèce *nilotica* (56.4 %) qu'au sein de la sous-espèce *paradoxa* (45.9 %), et inversement pour les proportions d'acide stéarique (avec respectivement : 29.7 % et 40.8 %) (**Tableau III.14**). Cette différenciation est également forte pour les teneurs en γ -tocophérol avec des teneurs plus importantes pour la sous-espèce *nilotica* (33.7 $\mu\text{g/g}$) par rapport à la sous-espèce *paradoxa* (10.5 $\mu\text{g/g}$) (**Tableau III.22**).



En revanche, contrairement aux hypothèses formulées par Maranz *et al.* et celle testée lors de ce travail de thèse (**HYPOTHÈSE 2**) sur les relations entre la température et la composition chimique des graines (Maranz & Wiesman, 2004; Maranz *et al.*, 2004), nous n'avons pas pu mettre en évidence de corrélations significatives entre l'évolution des composantes climatiques et édaphiques, et la variation des constituants chimiques des graines (**Tableaux III.19, III.20, III.27 et III.28**).

Si les résultats obtenus au cours ne permettent pas de conclure sur les déterminants de la variation chimique des graines de Karité, nous suggérons que celle-ci pourrait être liée à des événements adaptatifs, locaux et spécifiques à des populations particulières et non en relation avec un gradient adaptatif suivant un cline Nord-Sud.

Cette étude, enfin, aura permis le développement d'un outil haut débit de caractérisation des profils d'acides gras des graines de Karité par utilisation de la spectroscopie proche infrarouge (SPIR) (Davrieux *et al.*, 2010). La mise au point de cette nouvelle technologie pour le karité devrait fortement faciliter les analyses de la variation des constituants chimiques de graines sur de grands échantillons et améliorer nos connaissances sur cette question.

1.3 Étude de la variabilité des gènes candidats

Rappel des hypothèses testées (3 et 4) :

Le polymorphisme des gènes associés aux constituants chimiques est marqué par l'impact de la sélection naturelle (et/ou artificielle liée à la domestication).

Le polymorphisme des gènes impliqués dans les voies de biosynthèse est associé à la variation des constituants chimiques des graines.

Au cours de l'étude du polymorphisme nucléotidique de gènes candidats appartenant aux voies de biosynthèse des acides gras et des tocophérols, nous avons mis en évidence la duplication du gène *SAD* chez le Karité. De plus, nous observons que cette duplication entraîne une différenciation fondamentale entre les deux sous-espèces. En outre, l'une des copies (*SAD2*) apparaît comme fonctionnelle au sein de la sous-espèce *nilotica* et non-fonctionnelle au sein de la sous-espèce *paradoxa* (**Figure IV.8 et IV.9**). Nous suggérons ainsi que le doublement fonctionnel de ce gène chez les individus *nilotica* pourrait

être un élément explicatif fort de la différence observée sur le plan chimique entre les deux sous-espèces, avec des proportions accrues en acides gras insaturés chez *nilotica*. Nous notons également que cette différenciation fonctionnelle de *SAD2* entraîne une dynamique évolutive différente, et nous proposons en particulier que *SAD2* pourrait devenir un pseudogène au sein de *paradoxa*, ou un nouveau membre de la famille des acyl-acyl-désaturases au sein de *nilotica*. Nous proposons ainsi que ce gène pourrait constituer le miroir d'une possible spéciation en cours entre *nilotica* et *paradoxa*.

Les patrons de variation des gènes candidats (*SAD1*, *SAD2* et *VTE2-2*) suggèrent en Afrique de l'Ouest des effets de sélection positive ou l'impact d'une expansion démographique des populations (**Tableaux IV.4 et IV.6** ; et **Figures IV.12, IV.13 et IV.24**) (**HYPOTHÈSE 3**). Nous ne parvenons cependant pas à conclure sur ces patrons mais proposons plusieurs hypothèses cohérentes avec les connaissances scientifiques actuelles et le modèle biologique étudié. Ainsi, outre l'effet possible d'une expansion démographique en Afrique occidentale, nous pensons qu'une sélection de *SAD1* et de *VTE2-2* est cohérente avec le rôle adaptatif de la composition des graines en acides gras et en tocophérols (Linder, 2000) (Chapitre III.1). Bien que le caractère non-fonctionnel de *SAD2*, n'apparaisse pas cohérent avec un événement de sélection positive, nous suggérons que cette copie pourrait avoir un rôle dans le « *silencing* » de *SAD1* via l'action d'ARN interférents (Rodin & Riggs, 2003). De plus, nous suggérons que ce mécanisme pourrait être un autre élément explicatif de la différenciation chimique observée entre les sous-espèces. Enfin, nous proposons également que les patrons de sélection positive puissent également être liés à un processus de domestication de *Vitellaria paradoxa* en Afrique de l'Ouest qui aurait débuté sur le plateau de Mossi au Burkina Faso il y a près de 1000 ans (Lovett & Haq, 2000a; Maranz & Wiesman, 2003).

Enfin, au cours de cette étude, l'analyse des tests de neutralités au sein de *VTE2-2* (**Tableau IV.8 et Figure IV.28**) et l'étude d'association entre les allèles du gène et les teneurs en α -tocophérols (**Tableau IV.11**) chez les individus du gradient Mali (**HYPOTHÈSE 4**), nous permettent de mettre en évidence une possible adaptation locale des individus au Nord de l'aire, dans le site de Sassambourou. Ce résultat est de plus renforcé par des flux de gènes continus au sein de ce gradient comme en témoigne le F_{ST} faible et non significatif observé (**Tableau IV.10**).

2 DISCUSSION

Compte tenu de la synthèse des résultats, en liaison avec les hypothèses de l'étude présentées ci-dessus, nous nous proposons de discuter les résultats obtenus de façon plus transversale qu'au sein de chaque partie.

2.1 La variation au sein de l'espèce et l'impact climatique

➤ Climat passé

En cohérence avec l'hypothèse de départ que nous avons formulée (**HYPOTHÈSE 1**) Nous avons vu au travers de l'étude phylogéographique, que les variations climatiques de l'ère quaternaire, avec son maximum glaciaire il y a environ 20000 ans, ont eu un impact très différent en Afrique de l'Ouest et dans le reste de l'aire naturelle du Karité. Ce contraste spatial semble notamment en grande partie à l'origine de la forte structure observée à l'échelle de l'aire naturelle de *Vitellaria paradoxa*. Lors du dernier maximum glaciaire, l'Afrique orientale se serait particulièrement refroidie et les précipitations auraient fortement diminué, induisant des incursions du Sahara de près de 400 km plus au Sud (Lioubimtseva *et al.*, 1998), et réduisant drastiquement l'aire de distribution théorique de l'espèce autour de refuges potentiels près du lac Tchad et en Ouganda (**Figure II.7**). À l'inverse, la distribution occidentale de *Vitellaria paradoxa* aurait été relativement épargnée par ces changements climatiques, en particulier du fait d'un effet tampon relatif à la mousson en provenance du golf de Guinée (Braconnot *et al.*, 2000). De plus, en accord avec les hypothèses de Maley sur l'expansion des écosystèmes de savane au dernier maximum glaciaire (Maley, 1989; Maley, 1991; Maley, 1996), il est probable que le karité n'ait pas été restreint à cette époque au sein de refuges, à l'inverse des espèces tropicales humides (Dauby *et al.*, 2010; Lowe *et al.*, 2010). Les résultats obtenus tranchent donc avec ceux de la première évaluation de la diversité génétique du karité sur son aire naturelle (Fontaine *et al.*, 2004) qui suggérait le rôle du « Dahomey Gap », un couloir climatique apparu à l'Holocène tardif (il y a environ 3000 ans) au Togo et au Bénin, en perturbant l'aire de distribution de la forêt tropicale en Afrique de l'Ouest (Maley, 1989; Maley, 1991; Maley, 1996). Malheureusement, nous n'avons pu étudier ici l'impact de cet événement climatique mais les travaux de Maley (1989, 1991, 1996) et les résultats obtenus lors de cette thèse nous

permettent de penser que l'impact des variations climatiques survenues à cette période n'a pas été très important. Cette hypothèse mériterait cependant d'être vérifiée, notamment par l'utilisation des modèles de prédiction des niches écologiques du karité à cette période.

Ainsi, le dernier maximum glaciaire pourrait être à l'origine de la différenciation observée sur les marqueurs neutres, créant un isolement géographique très important et réduisant drastiquement la diversité à l'Est de l'aire naturelle. Cette observation est ainsi également cohérente avec les patrons de variation des gènes candidats. Ainsi, on observe globalement une forte différenciation entre *nilotica* et *paradoxa* pour les gènes étudiés (**Figures IV.16, IV.17, IV.18, IV.19, IV.20, IV.21, IV.25 et IV.26**) qui pourrait résulter d'un isolement des peuplements. Plus encore, cette dérive (avec un possible soutien de l'adaptation) pourrait permettre d'expliquer les patrons observés pour le gène *SAD2*, non fonctionnel au sein de la sous-espèce *paradoxa* et fonctionnel au sein de la sous espèce *nilotica*.

➤ L'impact des variations climatiques le long des gradients

Les variations environnementales au sein de l'aire de distribution de *Vitellaria paradoxa* sont fortes avec des altitudes variant en moyenne entre 100 et 1600 m (Hall *et al.*, 1996), une pluviométrie entre 400 et 1400 mm d'eau par an, et des températures moyennes variant entre 21 et 29 °C. Au cours des analyses des variations chimiques des graines en relation avec des variables climatiques basées sur la température et la pluviométrie, nous avons mis en évidence que les corrélations étaient très faibles pour tous les constituants chimiques mesurés (**Tableaux III.19, III.20, III.27 et III.28**). Ainsi, nous avons suggéré que l'impact des variables climatiques globales est relativement faible sur la variabilité observée dans notre plan d'expérience, et que cette variabilité s'exprimait peut-être plus en relation avec des événements d'adaptation locale (voir point de discussion 2.2). Cette observation est donc en contradiction globale avec notre hypothèse de départ (**HYPOTHÈSE 2**) et les hypothèses de Maranz (Maranz & Wiesman, 2004; Maranz *et al.*, 2004). Si les calendriers de floraison (Hall *et al.*, 1996), ou la fructification (Lovett & Haq, 2000a) montrent eux, une variation clinale, les résultats observés pour les constituants chimiques des graines sont cohérents avec la variation locale observée pour certains caractères morphologiques chez l'espèce (Lovett & Haq, 2000b; Sanou *et al.*, 2005; Sanou *et al.*, 2006). Il est cependant très

probable que les variables climatiques utilisées pour décrire les contraintes abiotiques ne seraient pas suffisantes et que d'autres déterminants abiotiques (variations microclimatiques ; composition fine du sol ; disponibilité en eau du sol ; dénivellation du sol...) ou biotique (âge des arbres ; bactéries telluriques ; fumier naturel apporté ; plantes associées...) non étudiés lors de cette thèse seraient plus pertinents pour expliquer les patrons observés.

➤ Futur

Tenant compte des résultats obtenus et des hypothèses formulées sur l'impact des variations climatiques passées et présentes, nous pouvons donc nous interroger sur l'impact du changement climatique actuel sur la diversité et la distribution future de l'espèce. En outre, les prévisions réalistes prédisent un réchauffement global de 2 à 3 °C d'ici 100 ans en Afrique sub-sahélienne (Ipcc, 2007; Parry, 2007; Solomon, 2007) (**Figure DG.1**), les prévisions les plus pessimistes vont jusqu'à +7 °C. Tenant compte des spécificités de l'Afrique sub-sahélienne, nous pouvons donc penser que l'impact de ces changements sera plus rapide en Afrique Orientale qu'en Afrique de l'Ouest.

Dans ce contexte, et compte tenu de la faible diversité génétique observée au sein de la partie orientale de l'aire naturelle de l'espèce, nous proposons qu'un effort particulièrement important soit entrepris afin d'assurer le maintien de la sous-espèce *nilotica* en Afrique de l'Est. Ce constat est en particulier cohérent avec la décision de l'IUCN de placer cette espèce dans la liste rouge des espèces vulnérables (<http://www.iucnredlist.org/apps/redlist/details/37083/0>).

Cependant, prévoir l'impact du réchauffement actuel est complexe (Parmesan, 2006) car il est très soudain en comparaison des événements de réchauffement postglaciaires passés (Hulme *et al.*, 2001) et contrairement à nos hypothèses, les prévisions montrent un réchauffement très important en Afrique de l'Ouest (**Figure DG.1**).

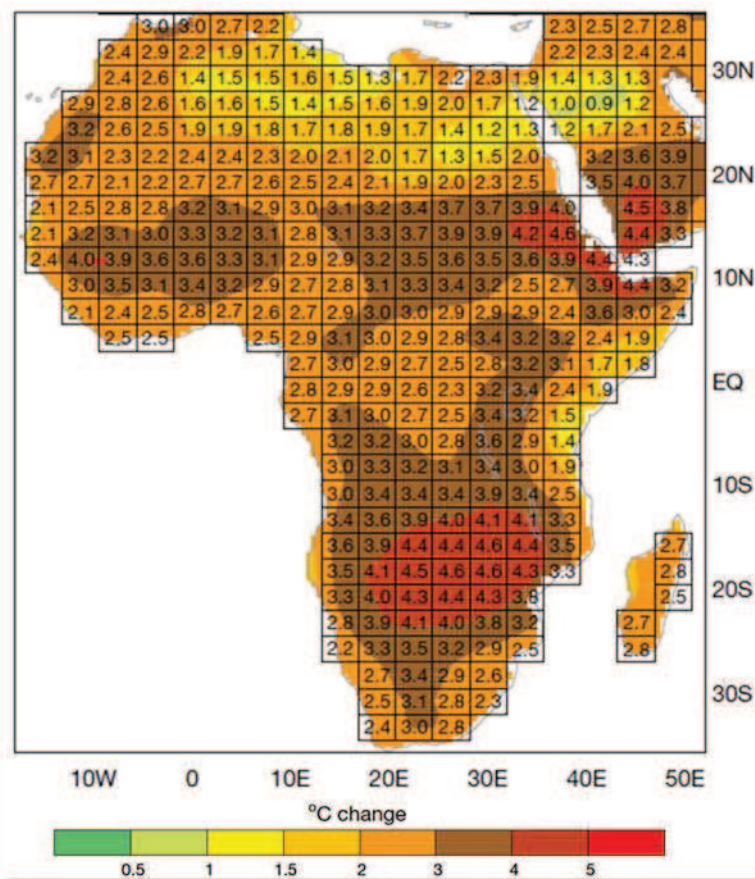


Figure DG.1. Prévisions du réchauffement climatique à horizon 2080 (Hulme *et al.*, 2001)

Lors d'un changement climatique, deux stratégies de survie sont possibles pour les êtres vivants : la migration vers un environnement favorable ou l'adaptation aux conditions nouvelles (Aitken *et al.*, 2008). Les arbres, organismes de taille et de longévité importante, ne dérogent pas à cette règle (Petit *et al.*, 2008). Malgré le symbole de grande stabilité que ces espèces représentent, des études ont notamment montré que leurs aires de distribution n'ont cessé d'évoluer en réponse aux contraintes climatiques passées (Bowen, 1979; Imbrie & Imbrie, 1980; Pias & Moore Jr, 1981; Davis & Shaw, 2001; Hamrick, 2004; Petit *et al.*, 2008). Bien que le changement climatique actuel soit très soudain, des études récentes ont mis en évidence que cette stratégie de migration de l'aire de distribution a débuté pour certaines espèces forestières (Davis & Zabinski, 1992; Davis & Shaw, 2001; Lloyd, 2005; Jump *et al.*, 2006). Dans ce contexte, il est cependant probable que certaines espèces végétales du biome de savane ne puissent changer leur distribution suffisamment rapidement (Woodward *et al.*, 2004). En particulier, le Karité est une espèce à croissance lente (Hall *et al.*, 1996) et arrivant tardivement à maturité reproductive (Chevalier, 1943; Ruysen, 1957; Bonkougou, 1987; Hall *et al.*, 1996). Il est donc possible que *Vitellaria paradoxa* ne puissent changer son aire de distribution rapidement en réponse à ce changement climatique.

Concernant la seconde option possible d'adaptation au climat changeant, des études récentes montrent que des espèces ont mis en place cette stratégie, en réponse au changement climatique actuel (Carter, 1996; Andalo *et al.*, 2005), adaptant en particulier leur croissance à ces nouvelles contraintes. Cependant, cette stratégie nécessite que l'espèce puisse faire preuve d'une plasticité phénotypique importante. Compte tenu de la vaste gamme de variation climatique dans laquelle le Karité évolue et de la rusticité de l'espèce, il est possible que *Vitellaria paradoxa* puisse s'adapter à ces nouvelles conditions environnementales. Cependant, de nouvelles l'évaluation de la plasticité phénotypique de cette espèce reste nécessaire, car, sur le plan des caractères chimiques et morphologiques étudiés au cours de cette thèse, nous n'avons mis en évidence que des coefficients de variation relativement faibles et une diversité génétique réduite en comparaison d'autres espèces d'arbre.

2.2 Impact de la sélection naturelle et humaine

➤ Par l'analyse des patrons de variation entre les sous-espèces

Dans cette thèse, nous avons mis en évidence à l'échelle de l'aire naturelle, des variations très importantes entre les deux sous-espèces *nilotica* et *paradoxa* (et entre les régions Afrique de l'Ouest et Afrique orientale) tant du point de vu chimique (proportions relatives des acides gras et teneurs en tocophérols) que sur le plan moléculaire (*SAD1*, *SAD2* et *VTE2-2*). Si notre plan d'expérimentation ne permet pas de distinguer les effets de la sélection des autres facteurs évolutifs (notamment la dérive) dans l'expression de cette variation, nous proposons ici quelques hypothèses.

L'Afrique de l'Ouest a un climat globalement plus chaud et plus sec que l'Afrique de l'Est (**Tableaux III.3** et **III.4**). Ces déterminants climatiques ont pu être à l'origine d'une adaptation à ces climats différents sous l'effet de la sélection naturelle. Cette adaptation pourrait ainsi être à l'origine de la proportion plus importante en acides gras insaturés en Afrique orientale, en réponse à l'environnement plus frais (Linder, 2000) (Chapitre III.1). Une hypothèse alternative pour expliquer ces teneurs différentes en acide gras serait l'impact d'une domestication. En effet, la tradition de l'utilisation du Karité par l'homme en agroforesterie est forte en Afrique de l'Ouest. Cette domestication aurait environ 1000 ans d'histoire au cœur de la région du plateau de Mossi au Burkina Faso (Lovett & Haq, 2000a;

Maranz & Wiesman, 2003). Les agriculteurs de cette région opèrent sur cet arbre une sélection consciente et inconsciente (Lovett & Haq, 2000a) dont les impacts positifs sur le plan morphologique ont notamment été rapportés par Lovett et Kelly (Lovett & Haq, 2000a; Kelly *et al.*, 2004; Kelly, 2005; Kelly *et al.*, 2007). Or, la préférence des femmes est portée vers l'obtention d'un beurre « solide », plus pratique que l'huile « liquide » pour le transport, le commerce et le stockage de ce produit dans cette région (Greenwood, 1929; Schreckenber, 1996; Lovett & Haq, 2000a). Dans ce contexte il est donc possible qu'une sélection anthropique des arbres produisant des graines à forte teneurs en acides saturés ait été effectuée dans cette région de grande tradition d'exploitation de l'espèce. Cette domestication en Afrique de l'Ouest pourrait ainsi être à l'origine de la signature de sélection positive enregistrée dans cette région pour les gènes étudiés. Cependant, comme nous l'avons déjà souligné, il est également possible que cette différenciation phénotypique et les empreintes moléculaires sur les gènes candidats soit une conséquence des phénomènes démographiques (expansion – **Box II.1**). Toutefois, cette expansion pourrait également résulter de l'impact positif des activités humaines traditionnelles éliminant les espèces ligneuses et non-ligneuses concurrentes, fertilisant les sols, commerçant ses graines... (Hall *et al.*, 1996; Lovett & Haq, 2000a; Kelly *et al.*, 2004; Kelly, 2005; Kelly *et al.*, 2007).

➤ Par l'analyse des patrons de variations selon le gradient Nord-Sud

À une échelle plus locale, l'analyse de quatre sites répartis sur un cline Nord-Sud au Mali, suggère une sélection positive au sein du gène *VTE2-2* associée à des teneurs accrues en α -tocophérol à l'extrême Nord de l'aire de distribution malienne au site de Sassambourou (**Tableau IV.8** et **Figure IV.28**). Au sein de ce gradient, la faible différenciation observée sur les marqueurs microsatellites (neutres) confirme qu'à cette échelle, l'impact des facteurs favorisant la structure tels que la dérive, est faible par comparaison avec l'échelle de l'aire naturelle. Ainsi, si cette adaptation locale nécessite d'être confirmée, elle pourrait résulter de l'effet de la sélection naturelle, car ce site se situe à la marge de l'aire naturelle, tant sur le plan géographique que climatique (**Tableau III.3** et **III.4**). Des conditions climatiques difficiles induisent un stress oxydatif accru chez les plantes (Brigelius-Flohé, 2009) pour lesquelles le maintien d'un mécanisme d'action antioxydante auxquels participent les tocophérols (et en particulier la forme α) devient une priorité afin d'éviter la peroxydation des lipides (Kamal-Eldin & Appelqvist, 1996; Brigelius-Flohé, 2009). Ainsi, la sélection positive

d'arbres plus résistants à ce climat chaud et sec pourrait être à l'origine des patrons de variation nucléotidique observés. Cependant, les activités humaines peuvent également être à l'origine de cette sélection locale. Lors de stress oxydatifs, les génotypes déficients pour l'enzyme homogénisate phytyltransférase codée par *VTE2*, caractérisés par des teneurs inférieures en tocophérols, montrent des phénotypes chétifs (Collakova & Dellapenna, 2003b, a). Or, nous avons souligné dans l'étude bibliographique (Chapitre I), que la sélection par élimination des arbres, opérée par les agriculteurs, se base notamment sur les caractéristiques de morphologie des arbres (Lovett & Haq, 2000a; Kelly, 2005). Ainsi, si l'on suppose que les arbres produisant des graines avec des teneurs supérieures en α -tocophérols montrent un développement plus important, il est possible que ces individus aient été favorisés lors de la gestion des jachères avant leur exploitation. Ces résultats sont cependant à relativiser du fait, d'une part de l'échantillonnage limité et, d'autre part, de la puissance du test qui devient non-significatif si on applique une correction de Bonferroni.

3 CONCLUSION ET PERSPECTIVES

3.1 Confirmation de résultats connus

De facto, lors de ce travail de thèse, nous avons confirmé certains résultats. C'est notamment le cas des patrons de variation chimiques des graines du karité (Maranz & Wiesman, 2003; Maranz *et al.*, 2003; Maranz & Wiesman, 2004; Maranz *et al.*, 2004; Di Vincenzo *et al.*, 2005) au travers desquels nous confirmons une **différenciation chimique forte** entre les sous-espèces. Cette **différenciation** est également vérifiée sur le plan **génétique**, en accord avec les premiers résultats fournis par Fontaine *et al.* (2004).

3.2 Nouveaux résultats et nouvelles hypothèses

Concernant les patrons de variations chimiques, nous ne confirmons pas l'hypothèse selon laquelle la température serait le facteur explicatif majeur de la variation observée (Maranz & Wiesman, 2004; Maranz *et al.*, 2004). Nos travaux suggèrent que la variation observée entre *nilotica* et *paradoxa* est liée à divers **effets combinés : dérive ; sélection** (naturelle et/ artificielle). Au regard des gradients climatiques étudiés, nous suggérons que certaines différences observées localement entre sites pourraient résulter d'une **adaptation locale**.

Nous proposons dans cette thèse que la **duplication du gène SAD** au sein de *Vitellaria paradoxa*, puisse être à l'**origine de la différenciation chimique** et puisse refléter une **spéciation allopatrique en cours** entre *nilotica* et *paradoxa*, résultant soit d'une **adaptation** aux aires de distributions non chevauchantes de ces sous-espèces, soit de l'effet de la **dérive génétique** liée à l'isolement géographique, soit de l'effet d'une **sélection anthropique**.

Nous proposons l'hypothèse que l'impact contrasté du refroidissement observé lors du **dernier maximum glaciaire** pourrait être à l'origine de la différenciation entre les sous-espèces *paradoxa* et *nilotica*, entraînant une forte **fragmentation** à l'Est avec le maintien de petites **zones refuges** (Figure II.7). Enfin nous proposons une **expansion démographique en Afrique de l'Ouest** qui pourrait être liée au maintien d'une zone favorable au **dernier maximum glaciaire** (il y a 20000 ans) ou à la **domestication** de l'espèce (il y a 1000 ans).

3.3 Perspectives de recherches

Afin de confirmer certains résultats obtenus et vérifier les hypothèses proposées, nous suggérons ici quelques perspectives de recherche.

Afin de mieux comprendre la variation phénotypique au sein de l'aire, sans attendre la mise en place d'essais multi-locaux (voir ci-après), il apparaît avantageux de renforcer l'échantillonnage au sein de l'aire. Notre étude, basée sur 600 individus, représentant un progrès significatif par rapport à la base de données de ProKarité (Prokarité_ [Http://Www.Prokarite.Org/](http://www.Prokarite.Org/); Maranz & Wiesman, 2003; Maranz *et al.*, 2003; Maranz & Wiesman, 2004; Maranz *et al.*, 2004; Di Vincenzo *et al.*, 2005), reste cependant très partielle compte tenu de l'aire de distribution de l'espèce et des multiples facteurs contribuant à la variabilité intra et inter populations. Les outils SPIR (spectroscopie proche infrarouge) élaborés dans cette thèse devraient largement contribuer à renforcer ces connaissances.

Afin de pousser plus avant les déterminants de la variation chimique des constituants de graines de karité et mieux appréhender la différenciation chimique observée entre les deux sous-espèces, il serait intéressant de mettre en place un plan expérimental en milieu contrôlé (dispositif impliquant des provenances et des descendances, implantées dans plusieurs lieux à l'Est et à l'Ouest). Ce dispositif permettrait de tester l'interaction génotype

X environnement pour dissocier les effets de plasticité phénotypique des effets portés par les gènes. Ce type de test requiert cependant une infrastructure et un plan expérimental difficile à mettre en place pour le Karité car les attributs biologiques (croissance lente, fructification tardive – chapitre I.2) constituent un frein important à la mise en place de ce protocole.

Concernant l'adaptation locale détectée au Nord du Mali pour les teneurs en α -tocophérols, nous proposons de tester cette hypothèse en utilisant : un échantillonnage plus important (nombre d'individus ; taille du fragment *VTE2-2* ; autres gènes d'intérêts tels que *VTE2-1* ou *TC – Tocopherol Cyclase*) ; des mesures répétées des teneurs en tocophérols sur plusieurs années pour les mêmes arbres ; des tests de neutralité sélective moins sensibles à la structure, comme le test HKA (Hudson *et al.*, 1987) ; et des modèles d'association mixtes avec les matrices d'apparementement entre arbres et/ou les variations climatiques comme cofacteurs (Yu *et al.*, 2006; Bradbury *et al.*, 2007; Kang *et al.*, 2008; Zhang *et al.*, 2010).

Concernant la duplication du gène *SAD*, de nombreuses questions restent à éclaircir. Sur le plan fonctionnel, nous proposons d'étudier l'impact physiologique de cette duplication et de la fonctionnalité/non-fonctionnalité de la copie *SAD2*. Pour cela il serait intéressant d'effectuer des dosages de transcrits *SAD1* et *SAD2*, et de la protéine *SAD* sur les plantules *nilotica* et *paradoxa* dont nous disposons en serre. Il serait également intéressant d'observer le gain phénotypique de lignées *Arabidopsis thaliana* mutantes *ssi2* (avec un gène *SAD* éteint) (Kachroo *et al.*, 2003), transformées ou non avec les copies *SAD1* et *SAD2* (fonctionnelle ou non). Afin de mieux appréhender la dynamique évolutive des deux copies de gène, nous envisageons une étude plus approfondie de la variation nucléotidique du gène, avec notamment un échantillonnage couvrant mieux la distribution des deux sous-espèces, pour lesquelles nous avons exprimé un doute taxonomique au cours du Chapitre II de cette thèse, au regard des patrons de différenciation observés.

Nous suggérons l'hypothèse d'une éventuelle spéciation en cours entre les sous-espèces. Comme nous l'avons souligné dans le Chapitre IV de cette thèse, il serait donc très intéressant d'une part de mettre en place des essais de fécondation croisée *nilotica* x *paradoxa*, et d'autre part d'étudier les patrons de diversité des marqueurs « classiques » des

processus de spéciations que sont les gènes, associés à l'isolement reproductif des populations (Rieseberg & Blackman, 2010).

L'hypothèse de la présence à l'Est de zones refuges glaciaires, à partir desquelles le karité aurait recolonisé le milieu, nécessite d'être confirmée. En particulier, il serait intéressant de comparer les résultats de modélisation des niches écologiques, obtenus par l'algorithme BIOCLIM (Booth *et al.*, 1987) de DIVA-GIS (Hijmans *et al.*, 2001) avec ceux proposés par d'autres logiciels utilisant d'autres modèles tels que MAXENT (Phillips *et al.*, 2004; Elith *et al.*, 2006; Phillips *et al.*, 2006). Afin de vérifier l'impact de ces refuges potentiels et le patron phylogéographique observé, il serait intéressant d'augmenter : le nombre d'individus et le nombre de sites échantillonnés à l'Est du Massif Adamawa, dans le but de rééquilibrer l'échantillon qui est plus dense à l'Ouest ; et le nombre de marqueurs microsatellites chloroplastiques, car bien que nous en ayons testés 10, seuls trois se sont avérés polymorphes chez *Vitellaria paradoxa* (Chapitre II).

S'agissant de l'expansion démographique détectée en Afrique de l'Ouest, il serait intéressant de confirmer cette hypothèse à travers une étude plus approfondie de la diversité génétique à l'échelle de l'Afrique de l'Ouest, avec en particulier, un nombre de sites et d'arbres échantillonnés plus important. Cette question fait l'objet du travail de thèse de Zénor Logossa sur un échantillonnage de 38 sites pour un total de 790 arbres, qui montre, en plus des résultats par l'approche d'ABC (voir chapitre II, Box I), de nombreux indices d'expansion, notamment au travers des patrons de « mismatch distributions » de marqueurs microsatellites chloroplastiques (Rogers & Harpending, 1992; Schneider & Excoffier, 1999; Excoffier, 2004; Excoffier *et al.*, 2005). Ce travail, en outre, fait l'objet d'une publication en collaboration en cours de revue à *Molecular Ecology* (Logossa *et al.* 2010). Nous envisageons enfin d'effectuer une démarche similaire concernant la partie orientale de l'aire naturelle, avec le test de scénarios bottleneck/expansion et de fusion des populations d'Afrique centrale et d'Afrique de l'Est.

Enfin, les méthodes de séquençage de nouvelle génération (NGS – « next-generation sequencing ») permettent d'envisager le développement de stratégies à plus large échelle pour répondre mieux aux différentes questions de recherches posées. Ces NGS, telles que les méthodes 454 (<http://www.454.com/>), Illumina (<http://www.illumina.com/>) ou encore

AB SOLiD (<http://www.appliedbiosystems.com>) permettent de séquencer des génomes entiers à moindre coût (Varshney *et al.*, 2009). Dans le cas de la mise en place d'une nouvelle étude d'association génotype/phénotype, les NGS, et, en particulier l'approche Illumina, offrent la possibilité de séquencer divers amplicons PCR d'un grand nombre de gènes candidats de plusieurs génotypes simultanément (Varshney *et al.*, 2009). Ainsi, en un simple « run » Illumina, nous pourrions générer un grand nombre de données de séquence pour un plus grand nombre de gènes candidats dans un délai et à un coût considérablement inférieurs à celui du séquençage traditionnel. De plus, ces NGS peuvent également permettre d'identifier un grand nombre de marqueurs sur le génome qui pourraient être ensuite utilisés pour les approches d'association en « whole genome scanning » (Rafalski, 2002; Nordborg & Weigel, 2008; Varshney *et al.*, 2009).

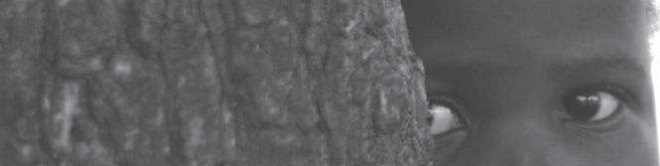
3.4 Perspectives de développement

Dans le contexte actuel d'un monde sous une forte dépendance de quelques espèces de plantes, la diversification des habitudes de production et de consommation afin d'inclure un plus large éventail d'espèces végétales, en particulier celles qui sont actuellement identifiées comme « sous-utilisées », peut contribuer de manière significative à améliorer la santé et la nutrition, les moyens de subsistance, la sécurité alimentaire des ménages et la durabilité écologique (Jaenicke & Höschle-Zeledon, 2006). Le karité appartient à ces espèces « sous-utilisées », offrant un énorme potentiel pour le développement durable, le maintien de la biodiversité indigène, des traditions locales et de la diversité culturelle (Jaenicke & Höschle-Zeledon, 2006). Pourtant, malgré un intérêt croissant pour ces espèces, Jaenicke et Höschle-Zeledon (2006) suggèrent que l'acquisition de connaissances plus approfondies concernant ces espèces « sous-utilisées » restent nécessaires sur le plan agronomique, économique et social, avec un focus particulier sur leur avenir face au réchauffement climatique actuel.

Ainsi, les résultats apportés par ce travail de thèse permettent de répondre partiellement à ces préoccupations, à travers le Karité. Ainsi, sur le plan scientifique, nous avons pu discuter quelques hypothèses larges sur l'impact des changements climatiques passés et de la domestication sur la variabilité de l'espèce. Cette base de réflexion que constitue ce document devrait donc permettre de mieux appréhender les questionnements



futurs sur le management de *Vitellaria paradoxa* au sein des parcs agroforestiers et sur la conservation de l'espèce. Sur le plan agro-économique, le développement d'une base de référence spectrale et d'équations de prédiction de la composition en acides gras des graines de Karité est une avancée également non négligeable qui devrait permettre à l'avenir une meilleure maîtrise de la qualité des beurres de Karité, répondant ainsi aux attentes des agriculteurs et des industriels (Davrieux *et al.*, 2010). Cependant, à l'instar de la majorité de ces espèces « sous-utilisées », nous constatons, à l'issue de cette thèse, que de nombreuses questions de recherche restent à éclaircir afin de percer les secrets de l'arbre « à beurre » d'Afrique.



Bibliographie

- Abbasi, A.-R., Hajirezaei, M., Hofius, D., Sonnewald, U. & Voll, L.M. (2007) Specific roles of α - and γ -tocopherol in abiotic stress responses of transgenic tobacco. *Plant Physiol.*, **143**, 1720-1738
- Abbiw, D. (1995) *Useful plants of ghana: West african uses of wild and cultivated plants*. Intermediate Technology Publications and The Royal Botanic Gardens, Londres, GB.
- Addlesee, H.A., Gibson, L.C., Jensen, P.E. & Hunter, C.N. (1996) Cloning, sequencing and functional assignment of the chlorophyll biosynthesis gene, chl_p, of synechocystis sp. Pcc 6803. *FEBS Lett*, **389**, 126-130
- Afnor (1997) Iso 9936:2006, animal and vegetable fats and oils – determination of tocopherol and tocotrienol contents by hplc. In. AFNOR, Paris
- Afzal-Rafii, Z. & Dodd, R.S. (2007) Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in western europe. *Molecular Ecology*, **16**, 723-736
- Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T. & Curtis-McLane, S. (2008) Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95-111
- Allal, F., Vaillant, A., Sanou, H., Kelly, B. & Bouvet, J.-M. (2008) Isolation and characterization of new microsatellite markers in shea tree (*Vitellaria paradoxa* c. F. Gaertn). *Molecular Ecology Resources*, **8**, 822-824
- Allen, J. (2000) *Middle egyptian: An introduction to the language and culture of hieroglyphs*. Cambridge Univ Pr.
- Amaral, J., Cunha, S., Alves, M., Pereira, J., Seabra, R. & Oliveira, B. (2004) Triacylglycerol composition of walnut (*Juglans regia* L.) cultivars: Characterization by hplc-elsd and chemometrics. *J. Agric. Food Chem*, **52**, 7964-7969
- Amaral, J.S., Alves, M.R., Seabra, R.M. & Oliveira, B.P.P. (2005) Vitamin e composition of walnuts (*Juglans regia* l.): A 3-year comparative study of different cultivars. *Journal of Agricultural and Food Chemistry*, **53**, 5467-5472

- Andalo, C., Beaulieu, J. & Bousquet, J. (2005) The impact of climate change on growth of local white spruce populations in québec, canada. *Forest Ecology and Management*, **205**, 169-182
- Anderberg, A.A., Rydin, C. & Kallersjo, M. (2002) Phylogenetic relationships in the order Ericales S.L.: Analyses of molecular data from five genes from the plastid and mitochondrial genomes. *Am. J. Bot.*, **89**, 677-687
- Anderson, D., Millar, J. & Weiping, W. (1991) Gum arabic (*Acacia senegal*) from Niger - comparison with other sources and potential agroforestry development. *Biochemical Systematics and Ecology*, **19**, 447-452
- Andolfatto, P., Wall, J.D. & Kreitman, M. (1999) Unusual haplotype structure at the proximal breakpoint of in(2l)t in a natural population of *Drosophila melanogaster*. *Genetics*, **153**, 1297-1311
- Arango, Y. & Heise, K.P. (1998) Tocopherol synthesis from homogentisate in *Capsicum annuum* L. (yellow pepper) chromoplast membranes: Evidence for tocopherol cyclase. *Biochem J*, **336**, 531-533
- Archer, S. (1996) Assessing and interpreting grass-woody plant dynamics. *The ecology and management of grazing systems* (ed. by J.H.a.A. Illius), pp. 101-134. CAB International, Wallingford, Oxon, United Kingdom.
- Arends, J. (1976) Somatic chromosome numbers of some African Sapotaceae. *Acta Bot. Neerl*, **25**, 449-457
- Aruna, R. & Nigam, S. (2009) Inheritance of fatty acid content and related quality traits in groundnut, *arachis hypogaea* L. *J. Oilseed Res.*, **26**(1), 10-17
- Ashman, T.-L., Knight, T.M., Steets, J.A., Amarasekare, P., Burd, M., Campbell, D.R., Dudash, M.R., Johnston, M.O., Mazer, S.J., Mitchell, R.J., Morgan, M.T. & Wilson, W.G. (2004) Pollen limitation of plant reproduction: Ecological and evolutionary causes and consequences. *Ecology*, **85**, 2408-2421
- Aubréville, A. (1950) *Flore forestière soudano-guinéenne*. Géogr., Marit. et Col., Paris.

- Bachlava, E., Auclair, J., Burton, J. & Cardinal, A. (2007) Heritability of oleic and linolenic acid seed content and their genetic correlations with quality and agronomic traits in soybean.
- Backéus, I. (1992) Distribution and vegetation dynamics of humid savannas in Africa and Asia. *Journal of Vegetation Science*, **3**, 345-356
- Bagley, M.J., Anderson, S.L. & May, B. (2001) Choice of methodology for assessing genetic impacts of environmental stressors: Polymorphism and reproducibility of RAPD and AFLP fingerprints. *Ecotoxicology*, **10**, 239-244
- Bailey, A.E. (1982) *Bailey's industrial oil and fat products*, 4th edn. Swern, D. (ed) John Wiley and Sons, Inc, New York, NY.
- Baker, H. (1962) Comments on the thesis that there was a major centre of plant domestication near the headwaters of the river niger. *The Journal of African History*, **3**, 229-233
- Balakirev, E.S. & Ayala, F.J. (2003) Pseudogenes: Are they “junk” or functional DNA? *Annual Review of Genetics*, **37**, 123-151
- Ballouche, A. & Neumann, K. (1995) A new contribution to the holocene vegetation history of the West-African sahel - pollen from oursi, Burkina-Faso and charcoal from 3 sites in northeast Nigeria. *Vegetation History and Archaeobotany*, **4**, 31-39
- Bandelt, H.J., Forster, P. & Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37-48
- Baumer, M. (1990) The potential role of agroforestry in combating desertification and environmental degradation with special reference to Africa. *Wageningen.: Technical Centre for Agricultural & Rural Cooperation 250 pp. Includes: Desertification and countries involved*
- Bayala, J., Mando, A., Teklehaimanot, Z. & Ouedraogo, S.J. (2005) Nutrient release from decomposing leaf mulches of Karite, (*Vitellaria paradoxa*) and Nere (*Parkia biglobosa*) under semi-arid conditions in burkina faso, west africa. *Soil Biology & Biochemistry*, **37**, 533-539

- Beaumont, M.A., Zhang, W. & Balding, D.J. (2002) Approximate bayesian computation in population genetics. *Genetics*, **162**, 2025-2035
- Beerling, D.J. & Osborne, C.P. (2006) The origin of the savanna biome. *Global Change Biology*, **12**, 2023-2031
- Behling, H. (2002) South and southeast Brazilian grasslands during late Quaternary times: A synthesis. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **177**, 19-27
- Beisswanger, S. & Stephan, W. (2008) Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in *Drosophila*. *Proceedings of the National Academy of Sciences, USA*, **105**, 5447
- Bekessy, S.A., Allnutt, T.R., Premoli, A.C., Lara, A., Ennos, R.A., Burgman, M.A., Cortes, M. & Newton, A.C. (2002) Genetic variation in the vulnerable and endemic monkey puzzle tree, detected using rapds. *Heredity*, **88**, 243-249
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. (2004) Genetix 4.05, logiciel sous windows tm pour la génétique des populations. Laboratoire génome, populations, interactions, cnrs umr 5171. *Université de Montpellier II, Montpellier (France)*
- Berhaut, J. (1967) *Flore du Sénégal*, 2 edn. Clairafrique, Dakar.
- Bessegga, C., Saidman, B.O., Darquier, M.R., Ewens, M., Sanchez, L., Rozenberg, P. & Vilardi, J.C. (2009) Consistency between marker- and genealogy-based heritability estimates in an experimental stand of *Prosopis alba* (Leguminosae). *Am. J. Bot.*, **96**, 458-465
- Bikard, D., Patel, D., Le Mette, C., Giorgi, V., Camilleri, C., Bennett, M.J. & Loudet, O. (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *Arabidopsis thaliana*. *Science*, **323**, 623-626
- Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, **72**, 291-336

- Blumler, M. (1996) Ecology, evolutionary theory and agricultural origins. *The origins and spread of agriculture and pastoralism in eurasia* (ed. by D.R. Harris), pp. 25-50. Smithsonian Inst Press, Londres, GB.
- Boege, K. & Marquis, R.J. (2005) Facing herbivory as you grow up: The ontogeny of resistance in plants. *Trends in Ecology & Evolution*, **20**, 441-448
- Böhle, U.R., Hilger, H.H. & Martin, W.F. (1996) Island colonization and evolution of the insular woody habit in *Echium* L. (*boraginaceae*). *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 11740-11745
- Bokdam, J. (1977) Seedling morphology of some African Sapotaceae and its taxonomical significance. *Meded. Landbouwhoges. Wageningen*, **77**, 1-84
- Bonaccorso, E., Koch, I. & Peterson, A.T. (2006) Pleistocene fragmentation of Amazon species' ranges. *Diversity and Distributions*, **12**, 157-164
- Bonkougou, E. (1987) *Monographie du karité, Butyrospermum paradoxum (gaertn. F.) hepper, espèce agroforestière à usages multiples*. Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, Ouagadougou, Burkina Faso.
- Booth, F. & Wickens, G. (1988) *Non-timber uses of selected arid zone trees and shrubs in africa*. Food & Agriculture Organization of the UN (FAO)
- Booth, T.H., Nix, H.A., Hutchinson, M.F. & Busby, J.R. (1987) Grid matching: A new method for homoclimate analysis. *Agricultural and Forest Meteorology*, **39**, 241-255
- Boring, L.R., Swank, W.T., Waide, J.B. & Henderson, G.S. (1988) Sources, fates, and impacts of nitrogen inputs to terrestrial ecosystems: Review and synthesis. *Biogeochemistry*, **6**, 119-159
- Bottin, L., Isnard, C., Lagrange, A. & Bouvet, J.M. (2007) Comparative molecular and phytochemical study of the tree species *Santalum austrocaledonicum* (Santalaceae) distributed in the new-caledonian archipelago. *Chem Biodivers*, **4**, 1541-1556
- Bouvet, J., Kelly, B., Sanou, H. & Allal, F. (2008) Comparison of marker- and pedigree-based methods for estimating heritability in an agroforestry population of *Vitellaria*



- paradoxa* c.F. Gaertn. (shea tree). *Genetic Resources and Crop Evolution*, **55**, 1291-1301
- Bouvet, J.M., Vigneron, P. & Saya, A. (2005) Phenotypic plasticity of growth trajectory and ontogenic allometry in response to density for eucalyptus hybrid clones and families. *Ann Bot (Lond)*, **96**, 811-821
- Bowen, D.Q. (1979) Geographical perspective on the quaternary. *Progress in Physical Geography*, **3**, 167-186
- Braconnot, P., Joussaume, S., De Noblet, N. & Ramstein, G. (2000) Mid-holocene and last glacial maximum african monsoon changes as simulated within the paleoclimate modelling intercomparison project. *Global and Planetary Change*, **26**, 51-66
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. & Buckler, E.S. (2007) Tassel: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633-2635
- Bradshaw, A.D. (1972) Some of the evolutionary consequences of being a plant. *Journal of Evolutionary Biology*, **5**, 25-47
- Bridges, C.B. (1936) The bar "gene" a duplication. *Science*, **83**, 210-211
- Brigelius-Flohé, R. (2009) Vitamin e: The shrew waiting to be tamed. *Free Radical Biology and Medicine*, **46**, 543-554
- Brown, G.R., Gill, G.P., Kuntz, R.J., Langley, C.H. & Neale, D.B. (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc Natl Acad Sci USA*, **101**, 15255-15260
- Bryan, G.J., McNicoll, J., Ramsay, G., Meyer, R.C. & De Jong, W.S. (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of solanaceous plants. *TAG Theoretical and Applied Genetics*, **99**, 859-867
- Busby, J.R. (1991) Bioclim - a bioclimate analysis and prediction system. *Nature conservation: Cost effective biological surveys and data analysis* (ed. by C.R.a.A. Margules, M.P.). CSIRO, Melbourne.

- Cahoon, E.B., Becker, C.K., Shanklin, J. & Ohlrogge, J.B. (1994) Cdnas for isoforms of the delta 9-stearoyl-acyl carrier protein desaturase from *Thunbergia alata* endosperm. *Plant Physiol*, **106**, 807-808
- Cahoon, E.B., Mills, L.A. & Shanklin, J. (1996) Modification of the fatty acid composition of *Escherichia coli* by coexpression of a plant acyl-acyl carrier protein desaturase and ferredoxin. *Journal of Bacteriology*, **178**, 936-939
- Cahoon, E.B., Shah, S., Shanklin, J. & Browse, J. (1998) A determinant of substrate specificity predicted from the acyl-acyl carrier protein desaturase of developing cat's claw seed. *Plant Physiol.*, **117**, 593-598
- Calvo, R. & Horvitz, C. (1990) Pollinator limitation, cost of reproduction, and fitness in plants: A transition-matrix demographic approach. *American Naturalist*, **136**, 499-516
- Camus-Kulandaivelu, L., Veyrieras, J.B., Madur, D., Combes, V., Fourmann, M., Barraud, S., Dubreuil, P., Gouesnard, B., Manicacci, D. & Charcosset, A. (2006) Maize adaptation to temperate climate: Relationship between population structure and polymorphism in the dwarf8 gene. *Genetics*, **172**, 2449-2463
- Cao, Y., Xian, M., Yang, J., Xu, X., Liu, W. & Li, L. (2010) Heterologous expression of stearoyl-acyl carrier protein desaturase (s-acp-des) from *Arabidopsis thaliana* in escherichia coli. *Protein Expression and Purification*, **69**, 209-214
- Carter, K. (1996) Provenance tests as indicators of growth response to climate change in 10 north temperate tree species. *Canadian Journal of Forest Research*, **26**, 1089-1095
- Casini, C., Dardanelli, J.L., Martinez, M.J., Balzarini, M., Borgogno, C.S. & Nassetta, M. (2003) Oil quality and sugar content of peanuts (*Arachis hypogaea*) grown in Argentina: Their relationship with climatic variables and seed yield. *J Agric Food Chem*, **51**, 6309-6313
- Cavalli-Sforza, L. (1967) Phylogenetic analysis: Models and estimation procedures. *Am J Hum Genet*, **19**, 233-257
- Cavalli-Sforza, L.L. (1966) Population structure and human evolution. *Proc R Soc Lond B Biol Sci*, **164**, 362-379

- Cerling, T.E., Harris, J.M., Macfadden, B.J., Leakey, M.G., Quade, J., Eisenmann, V. & Ehleringer, J.R. (1997) Global vegetation change through the miocene/pliocene boundary. *Nature*, **389**, 153-158
- Charlesworth, J. & Eyre-Walker, A. (2006) The rate of adaptive evolution in enteric bacteria. *Mol Biol Evol*, **23**, 1348-1356
- Chevalier, A. (1943) Le karité ou arbre à beurre: Essai monographique. *Rev. Int. Bot. Appl. Agric. Trop*, **23**, 100-120
- Chevalier, A. (1946) L'arbre à beurre d'Afrique et l'avenir de sa culture. *Oléagineux*, **1**, 7-11
- Chevalier, A. (1948) Nouvelles recherches sur l'arbre à beurre du soudan, *Butyrospermum paradoxum*. *Revue de Botanique Appliquée* **28**, **303**, 241-256
- Ching, L.S. & Mohamed, S. (2001) Alpha-tocopherol content in 62 edible tropical plants. *Journal of Agricultural and Food Chemistry*, **49**, 3101-3105
- Chipp, T. (1927) *The gold coast forest: A study in synecology*. The Clarendon Press, Oxford.
- Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W. & McCabe, A.M. (2009) The last glacial maximum. *Science*, **325**, 710-714
- Cockram, J., White, J., Leigh, F.J., Lea, V.J., Chiapparino, E., Laurie, D.A., Mackay, I.J., Powell, W. & O'sullivan, D.M. (2008) Association mapping of partitioning loci in barley. *BMC Genetics*, **9**, 16
- Collakova, E. & Dellapenna, D. (2001) Isolation and functional analysis of homogentisate phytyltransferase from *Synechocystis* sp. Pcc 6803 and *Arabidopsis*. *Plant Physiol*, **127**, 1113-1124
- Collakova, E. & Dellapenna, D. (2003a) Homogentisate phytyltransferase activity is limiting for tocopherol biosynthesis in *Arabidopsis*. *Plant Physiol*, **131**, 632-642
- Collakova, E. & Dellapenna, D. (2003b) The role of homogentisate phytyltransferase and other tocopherol pathway enzymes in the regulation of tocopherol synthesis during abiotic stress. *Plant Physiol*, **133**, 930-940

- Collin, V., Eymery, F., Genty, B., Rey, P. & Havaux, M. (2008) Vitamin e is essential for the tolerance of *Arabidopsis thaliana* to metal-induced oxidative stress. *Plant, Cell & Environment*, **31**, 244-257
- Collins, W.D., Blackmon, M.L., Bonan, G.B., Hack, J.J., Henderson, T.B., Kiehl, J.T., Large, W.G., McKenna, D.S., Bitz, C.M., Bretherton, C.S., Carton, J.A., Chang, P., Doney, S.C., Santer, B.D. & Smith, R.D. (2006) The community climate system model version 3 (ccsm3). *Journal of Climate*, **19**, 2122-2143
- Collinson, C. & Zedie-Bosuener, A. (1999) *Shea butter markets-implications for ghanaian shea butter processors and exporters. Report no. 2403*. Natural Resources Institute, University of Greenwich, UK.
- Comps, B., Gomory, D., Letouzey, J., Thiebaut, B. & Petit, R.J. (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, **157**, 389-397
- Cornuet, J.-M., Ravigne, V. & Estoup, A. (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software diyabc (v1.0). *BMC Bioinformatics*, **11**, 401
- Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T. & Estoup, A. (2008) Inferring population history with diy abc: A user-friendly approach to approximate bayesian computation. *Bioinformatics*, **24**, 2713-2719
- Coull, G.C. (1928) Distribution and yields of shea butter nut trees in the northern territories. *Yearbook of the gold coast*. Department of Agriculture, Accra.
- Cronquist, A. (1981) *An integrated system of classification of flowering plants*. Columbia University Press, New York.
- Csilléry, K., Blum, M.G.B., Gaggiotti, O.E. & François, O. (2010) Approximate bayesian computation (abc) in practice. *Trends in Ecology & Evolution*, **25**, 410-418

- D'harlingue, A. & Camara, B. (1985) Plastid enzymes of terpenoid biosynthesis. Purification and characterization of gamma-tocopherol methyltransferase from capsicum chromoplasts. *J Biol Chem*, **260**, 15200-15203
- Dahlgren, R. (1983) General aspects of angiosperm evolution and macrosystematics. *Nordic Journal of Botany*, **3**, 119-149
- Dauby, G., Duminil, J., Heuertz, M. & Hardy, O. (2010) Chloroplast DNA polymorphism and phylogeography of a central African tree species widespread in mature rainforests: *Greenwayodendron suaveolens* (Annonaceae). *Tropical Plant Biology*, **3**, 4-13
- Davis, M. & Zabinski, C. (1992) Changes in geographical range resulting from greenhouse warming: Effects on biodiversity in forests. *Global warming and biological diversity*, pp. 297–308. Yale University Press.
- Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to quaternary climate change. *Science*, **292**, 673-679
- Davrieux, F., Allal, F., Piombo, G., Kelly, B., Okulo, J.B., Thiam, M., Diallo, O.B. & Bouvet, J.-M. (2010b) Near infrared spectroscopy for high-throughput characterization of shea tree (*Vitellaria paradoxa*) nut fat profiles. *Journal of Agricultural and Food Chemistry*, **58**, 7811-7819
- Demesure, B., Sodji, N. & Petit, R.J. (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, **4**, 129-134
- Depaulis, F. & Veuille, M. (1998) Neutrality tests based on the distribution of haplotypes under an infinite-site model. *Molecular Biology and Evolution*, **15**, 1788-1790
- Deyholos, M.K. & Sieburth, L.E. (2000) Separable whorl-specific expression and negative regulation by enhancer elements within the agamous second intron. *Plant Cell*, **12**, 1799-1810
- Di Vincenzo, D., Maranz, S., Serraiocco, A., Vito, R., Wiesman, Z. & Bianchi, G. (2005) Regional variation in shea butter lipid and triterpene composition in four African countries. *Journal of Agricultural and Food Chemistry*, **53**, 7473-7479

- Dick, C. (2010) Phylogeography and population structure of tropical trees. *Tropical Plant Biology*, **3**, 1-3
- Dick, C.W. & Heuertz, M. (2008) The complex biogeographic history of a widespread tropical tree species. *Evolution*, **62**, 2760-2774
- Druitt, D., Enright, N. & Ogden, J. (1990) Altitudinal zonation in the mountain forests of mt hauhungatahi, north island, New Zealand. *Journal of Biogeography*, **17**, 205-220
- Dumolin-Lapegue, S., Pemonge, M.-H. & Petit, R.J. (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology*, **6**, 393-397
- Eckert, A.J., Wegrzyn, J.L., Pande, B., Jermstad, K.D., Lee, J.M., Liechty, J.D., Tearse, B.R., Krutovsky, K.V. & Neale, D.B. (2009) Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal douglas fir (*pseudotsuga menziesii* var. *Menziesii*). *Genetics*, **183**, 289-298
- Eggeling, W. (1951) *The indigenous trees of the uganda protectorate*, 2 edn. Govt. Printer, Entebbe.
- Ehleringer, J., Sage, R., Flanagan, L. & Pearcy, R. (1991) Climate change and the evolution of C4 photosynthesis. *Trends in Ecology & Evolution*, **6**, 95-99
- Ehret, C. (1984) Historical/linguistic evidence for early african food production. *From hunters to farmers: The causes and consequences of food production in Africa* (ed. by J.D.B. Clark, S.A.), pp. 26–36. University of California Press, Berkeley, CA.
- Ehrlén, J. & Lehtilä, K. (2002) How perennial are perennial plants? *Oikos*, **98**, 308-322
- El Ghazali, G. (1993) A study on the pollen flora of sudan. *Review of Palaeobotany and Palynology*, **76**, 99-345
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams,

- S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151
- Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, **14**, 2611-2620
- Excoffier, L. (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: Lessons from the infinite-island model. *Molecular Ecology*, **13**, 853-864
- Excoffier, L., Hofer, T. & Foll, M. (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285-298
- Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47-50
- Falk, J. & Munné-Bosch, S. (2010) Tocochromanol functions in plants: Antioxidation and beyond. *Journal of Experimental Botany*, **61**, 1549
- Farquhar, G., Ehleringer, J. & Hubick, K. (1989) Carbon isotope discrimination and photosynthesis. *Annual Review of Plant Biology*, **40**, 503-537
- Farris, J.S. (1970) Methods for computing wagner trees. *Systematic Zoology*, **19**, 83-92
- Fay, J.C. & Wu, C.I. (2000) Hitchhiking under positive darwinian selection. *Genetics*, **155**, 1405-1413
- Feller, C., Lavelle, P., Albrecht, A. & Nicolardot, B. (1993) La jachère et le fonctionnement des sols tropicaux : Rôle de l'activité biologique et des matières organiques : Quelques éléments de réflexion. *La jachère en afrique de l'ouest* (ed. by C. Floret and G. Serpantié), pp. 15-32. Orstom, Paris.
- Felsenstein, J. (1993) Phylip (phylogeny inference package) In. Department of Genetics, University of Washington, Seattle
- Fielding, A.H. & Bell, J.F. (1997) *A review of methods for the assessment of prediction errors in conservation presence/absence models*. Cambridge University Press, Cambridge, Royaume-Uni.



- Fiume, E., Christou, P., Giani, S. & Breviario, D. (2004) Introns are key regulatory elements of rice tubulin expression. *Planta*, **218**, 693-703
- Flagel, L.E. & Wendel, J.F. (2009) Gene duplication and evolutionary novelty in plants. *New Phytologist*, **183**, 557-564
- Flenley, J.R. (1998) Tropical forests under the climates of the last 30,000 years. *Climatic Change*, **39**, 177-197
- Floret, C. & Pontanier, R. (2000) *La jachère en Afrique tropicale: Rôles, aménagement, alternatives*. John Libbey Eurotext.
- Fofana, B., Cloutier, S., Duguid, S., Ching, J. & Rampitsch, C. (2006) Gene expression of stearoyl-acp desaturase and delta12 fatty acid desaturase 2 is modulated during seed development of flax (*linum usitatissimum*). *Lipids*, **41**, 705-712
- Fontaine, C., Lovett, P.N., Sanou, H., Maley, J. & Bouvet, J.M. (2004) Genetic diversity of the shea tree (*Vitellaria paradoxa* C.F. Gaertn.), detected by rapd and chloroplast microsatellite markers. *Heredity*, **93**, 639-648
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.-L. & Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**, 1531-1545
- Freeling, M. & Thomas, B.C. (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Research*, **16**, 805-814
- Friedman, R. & Hughes, A.L. (2001) Pattern and timing of gene duplication in animal genomes. *Genome Research*, **11**, 1842-1847
- Frohman, M.A., Dush, M.K. & Martin, G.R. (1988) Rapid production of full-length cdnas from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 8998-9002



- Frost, P. (1986) Responses of savannas to stress and disturbance: A proposal for a collaborative programme of research.
- Fu, Y.X. & Li, W.H. (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709
- Gallais, A. (1990) Théorie de la sélection en amélioration des plantes. *Collection sciences agronomiques*.
- Garcia, I., Rodgers, M., Lenne, C., Rolland, A., Sailland, A. & Matringe, M. (1997) Subcellular localization and purification of a p-hydroxyphenylpyruvate dioxygenase from cultured carrot cells and characterization of the corresponding cdna. *Biochemical Journal*, **325**, 761
- Garcia, I., Rodgers, M., Pepin, R., Hsieh, T.-F. & Matringe, M. (1999) Characterization and subcellular compartmentation of recombinant 4-hydroxyphenylpyruvate dioxygenase from arabidopsis in transgenic tobacco. *Plant Physiol.*, **119**, 1507-1516
- Garnier-Géré, P.H. & Ades, P.K. (2001) Environmental surrogates for predicting and conserving adaptive genetic variability in tree species. *Conservation Biology*, **15**, 1632-1644
- Garza, J.C. & Williamson, E.G. (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305-318
- Gilliland, L.U., Magallanes-Lundback, M., Hemming, C., Supplee, A., Koornneef, M., Bentsink, L. & Dellapenna, D. (2006) Genetic basis for natural variation in seed vitamin e levels in arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 18834-18841
- Gillis, M., Perkins, D.H., Roemer, M. & Snodgrass, D.R. (1998) *Économie du développement* 2e édition edn. Nouveaux Horizons, Bruxelles.
- Giombini, M., Frankel, N., Iusem, N. & Hasson, E. (2009) Nucleotide polymorphism in the drought responsive gene asr2 in wild populations of tomato. *Genetica*, **136**, 13-25
- Giovannoni, J.J. (2004) Genetic regulation of fruit development and ripening. *Plant Cell*, **16** Suppl, S170-80

- Goffman, F.D. & Becker, H.C. (2001a) Diallel analysis for tocopherol contents in seeds of rapeseed. *Crop Science*, **41**, 1072-1079
- Goffman, F.D. & Becker, H.C. (2001b) Genetic analysis of tocopherol content and composition in winter rapeseed. *Plant Breeding*, **120**, 182-184
- Goffman, F.D. & Mollers, C. (2000) Changes in tocopherol and plastochromanol-8 contents in seeds and oil of oilseed rape (*Brassica napus* L.) during storage as influenced by temperature and air oxygen. *J Agric Food Chem*, **48**, 1605-1609
- Gonzalez-Martinez, S.C., Alia, R. & Gil, L. (2002) Population genetic structure in a mediterranean pine (*Pinus pinaster* Ait.): A comparison of allozyme markers and quantitative traits. *Heredity*, **89**, 199-206
- Goudet, J. (1995) Fstat (version 1.2): A computer program to calculate f-statistics. *Journal of Heredity*, **86**, 485
- Goudet, J. (2001) Fstat, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www2.Unil.Ch/popgen/softwares/fstat.Htm>. In:
- Govaerts, R., Frodin, D. & Pennington, T. (2001) *World checklist and bibliography of sapotaceae*. Kew: Royal Botanic Gardens, Kew, Londres, GB.
- Greenwood, M. (1929) Shea nuts and shea butter. *Bulletin of the Agricultural Department*, 59-100
- Grivet, D., Heinze, B., Vendramin, G.G. & Petit, R.J. (2001) Genome walking with consensus primers: Application to the large single copy region of chloroplast DNA. *Molecular Ecology Notes*, **1**, 345-349
- Gros, A. (1979) *Guide pratique de la fertilisation*. La Maison Rustique, Paris.
- Gruenewald, H., Brandt, B., Schneider, B., Bens, O., Kendzia, G. & Hüttel, R. (2007) Agroforestry systems for the production of woody biomass for energy transformation purposes. *Ecological Engineering*, **29**, 319-328

- Grusak, M.A. & Dellapenna, D. (1999) Improving the nutrient composition of plants to enhance human nutrition and health. *Annual Review of Plant Physiology and Plant Molecular Biology*, **50**, 133-161
- Guignard, J., Cosson, L. & Henry, M. (1985) *Abrégé de phytochimie*. Masson.
- Gunstone, F., Harwood, J. & Padley, F. (1994) *The lipid handbook*. Chapman & Hall/CRC.
- Gura, T. (2000) A silence that speaks volumes. *Nature*, **404**, 804-808
- Gutschick, V. & Bassirad, H. (2003) Extreme events as shaping physiology, ecology, and evolution of plants: Toward a unified definition and evaluation of their consequences. *New Phytologist*, **160**, 21-42
- Hall, J.B., Aebischer, D.P., Tomlinson, H.F., Osei-Amaning, E. & Hindle, J.R. (1996) *Vitellaria paradoxa: A monograph*. School of Agricultural and Forest Sciences, University of Wales Bangor, UK.
- Hamblin, M.T., Mitchell, S.E., White, G.M., Gallego, J., Kukatla, R., Wing, R.A., Paterson, A.H. & Kresovich, S. (2004) Comparative population genetics of the panicoid grasses: Sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. *Genetics*, **167**, 471-483
- Hamilton, A.J. & Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, **286**, 950-952
- Hampe, A. & Bairlein, F. (2000) Modified dispersal-related traits in disjunct populations of bird-dispersed *Frangula alnus* (Rhamnaceae): A result of its quaternary distribution shifts? *Ecography*, **23**, 603-613
- Hamrick, J., Godt, M., Murawski, D. & Loveless, M. (1991) Correlations between species traits and allozyme diversity: Implications for conservation biology. *Genetics and conservation of rare plants*, pp. 75–86. Oxford University Press.
- Hamrick, J.L. (2004) Response of forest trees to global environmental changes. *Forest Ecology and Management*, **197**, 323-335

- Hamrick, J.L. & Godt, M.J.W. (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **351**, 1291-1298
- Hamrick, J.L., Godt, M.J.W. & Sherman-Broyles, S.L. (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95-124
- Hamrick, J.L., Murawski, D.A. & Nason, J.D. (1993) The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Plant Ecology*, **107-108**, 281-297
- Hanley, J. & McNeil, B. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29
- Hardy, O.J. & Vekemans, X. (2002) Spagedi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620
- Harlan, J. (1971) Agricultural origins: Centers and noncenters. *Science*, **174**, 468-474
- Hemsley, J. (1968) Sapotaceae. *Flora of tropical east africa* (ed. by E. Milne-Redhead and R.M. Polhill), p. 78. Crown Agents for Oversea Governments and Administrations, Londres.
- Herwitz, S., Slye, R. & Turton, S. (2000) Long-term survivorship and crown area dynamics of tropical rain forest canopy trees. *Ecology*, **81**, 585-597
- Heuertz, M., Carnevale, S., Fineschi, S., Sebastiani, F., Hausman, J.F., Paule, L. & Vendramin, G.G. (2006) Chloroplast DNA phylogeography of european ashes, *Fraxinus* sp. (oleaceae): Roles of hybridization and life history traits. *Mol Ecol*, **15**, 2131-2140
- Heuertz, M., Fineschi, S., Anzidei, M., Pastorelli, R., Salvini, D., Paule, L., Frascaria-Lacoste, N., Hardy, O.J., Vekemans, X. & Vendramin, G.G. (2004) Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in europe. *Mol Ecol*, **13**, 3437-3452



- Heuertz, M., Teufel, J., González-Martínez, S.C., Soto, A., Fady, B., Alía, R. & Vendramin, G.G. (2010) Geography determines genetic relationships between species of mountain pine (*Pinus mugo* complex) in western Europe. *Journal of Biogeography*, **37**, 541-556
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978
- Hijmans, R.J., Guarino, L., Cruz, M. & Rojas, E. (2001) Computer tools for spatial analysis of plant genetic resources data: 1. Diva-gis. *Plant Genetic Resources Newsletter*, **127**, 15-19
- Hill, W.G. & Robertson, A. (1968) Linkage disequilibrium in finite populations. *THEORETICAL AND APPLIED GENETICS*, **38**, 226-231
- Hoballah, M.E., Gubitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell'olivo, A., Arnold, M. & Kuhlemeier, C. (2007) Single gene-mediated shift in pollinator attraction in petunia. *Plant Cell*, **19**, 779-790
- Hoefsloot, H., Van Der Pol, F. & L, R. (1993) *Jachère améliorées : Option pour le développement des systèmes de production en afrique de l'ouest* Bull.333, Amsterdam.
- Holland, J. (1922) The useful plants of niger. *Bulletin of Miscellaneous Information*, **Ad**, **9**, 1-963
- Hou, G., Ablett, G., Pauls, K. & Rajcan, I. (2006) Environmental effects on fatty acid levels in soybean seed oil. *Journal of the American Oil Chemists' Society*, **83**, 759-763
- House, J.I., Archer, S., Breshears, D.D., Scholes, R.J. & Participants, N.T.G.I. (2003) Conundrums in mixed woody-herbaceous plant systems. *Journal of Biogeography*, **30**, 1763-1777
- Howe, G.T., Aitken, S.N., Neale, D.B., Jermstad, K.D., Wheeler, N.C. & Chen, T.H. (2003) From genotype to phenotype: Unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany*, **81**, 1247-1266

- Huang, A. (1996) Oleosins and oil bodies in seeds and other organs. *PLANT PHYSIOLOGY*, **110**, 1055
- Hudson, R.R., Bailey, K., Skarecky, D., Kwiatowski, J. & Ayala, F.J. (1994) Evidence for positive selection in the superoxide-dismutase (sod) region of *Drosophila melanogaster*. *Genetics*, **136**, 1329-1340
- Hudson, R.R., Kreitman, M. & Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153-159
- Hudson, R.R., Slatkin, M. & Maddison, W.P. (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583-589
- Hughes, A.L. (1994) The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **256**, 119-124
- Hulme, M., Doherty, R., Ngara, T., New, M. & Lister, D. (2001) African climate change: 1900-2100. *Climate Research*, **17**, 145-168
- Hurme, P., Repo, T., Savolainen, O. & Pääkkönen, T. (1997) Climatic adaptation of bud set and frost hardiness in scots pine (*Pinus sylvestris*). *Canadian Journal of Forest Research*, **27**, 716-723
- Hurst, L.D. (2009) Fundamental concepts in genetics: Genetics and the understanding of selection. *Nat Rev Genet*, **10**, 83-93
- Husband, B. & Schemske, D. (1996) Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution*, **50**, 54-70
- Hutchinson, J. (1973) *The families of flowering plants: Arranged according to a new system based on their probable phylogeny*. Oxford: Clarendon Press xx, 968p.. Illustrations, maps.
- Hutchinson, J. & Dalziel, J. (1973) *Flora of west tropical Africa*. Whitefriars Press, London, GB.
- Hyman, E. (1991) A comparison of labor-saving technologies for processing shea nut butter in mali. *World development*, **19**, 1247-1268

- Imbrie, J. & Imbrie, J.Z. (1980) Modeling the climatic response to orbital variations. *Science*, **207**, 943-953
- Ingvarsson, P.K. (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of european aspen (*Populus tremula* L., salicaceae). *Genetics*, **169**, 945-953
- Ingvarsson, P.K., Garcia, M.V., Luquez, V., Hall, D. & Jansson, S. (2008) Nucleotide polymorphism and phenotypic associations within and around the phytochrome b2 locus in european aspen (*Populus tremula*, salicaceae). *Genetics*, **178**, 2217-2226
- Ipcc (2007) Ipcc fourth assessment report on climate change. In: *The Physical Science Basis*
- Isleib, T., Pattee, H., Sanders, T., Hendrix, K. & Dean, L. (2006) Compositional and sensory comparisons between normal-and high-oleic peanuts. *J. Agric. Food Chem*, **54**, 1759-1763
- Jackson, G. (1969) Notes on west African vegetation-iii. The seedling morphology of *Butyrospermum paradoxum* (gaertn. F.) hepper. Jl. W. *Afr. Sci. Ass*, **13**, 215-22
- Jackson, G. (1974) Cryptogeal germination and other seedling adaptations to the burning of vegetation in savanna regions: The origin of the pyrophytic habit. *New Phytologist*, **73**, 771-780
- Jaenicke, H. & Höschle-Zeledon, I. (2006) *Strategic framework for underutilized plant species research and development: With special reference to asia and the pacific, and to sub-saharan africa*. Bioversity International.
- Jaramillo-Correa, J.P., Aguirre-Planter, E., Khasa, D.P., Eguiarte, L.E., Pinero, D., Furnier, G.R. & Bousquet, J. (2008) Ancestry and divergence of subtropical montane forest isolates: Molecular biogeography of the genus *Abies* (pinaceae) in southern mexico and guatemala. *Mol Ecol*, **17**, 2476-2490
- Johnsen, Ø. & Skrøppa, T. (2000) Provenances and families show different patterns of relationship between bud set and frost hardiness in *Picea abies*. *Canadian Journal of Forest Research*, **30**, 1858-1866

- Johnson, L.A. & Soltis, D.E. (1994) Matk DNA sequences and phylogenetic reconstruction in saxifragaceae s. Str. *Systematic Botany*, **19**, 143-156
- Johnson, M. (1991) Cytology. *The genera of sapotaceae* (ed. by T. Pennington), pp. 15-22. Royal Botanic Gardens Kew & New York Botanical Garden, New York.
- Johnston, B. & Berg, R. (1990) *Les stratégies gouvernementales en matière de développement agricole*. Economica.
- Jónsson, T.H. (2004) Stature of sub-arctic birch in relation to growth rate, lifespan and tree form. *Annals of Botany*, **94**, 753-762
- Joobeur, T., Viruel, M., De Vicente, M., Jauregui, B., Ballester, J., Dettori, M., Verde, I., Truco, M., Messeguer, R. & Batlle, I. (1998) Construction of a saturated linkage map for prunus using an almond × peach f2 progeny. *TAG Theoretical and Applied Genetics*, **97**, 1034-1041
- Jose, S. (2009) Agroforestry for ecosystem services and environmental benefits: An overview. *Agroforestry Systems*, **76**, 1-10
- Jouve, P. (1993) Usages et fonctions de la jachère en Afrique de l'ouest et au maghreb. *La jachère en afrique de l'ouest* (ed. by C. Floret and G. Serpantié), pp. 55-66. ORSTOM, Paris.
- Jump, A.S., Hunt, J.M., Pe & Uelas, J. (2006) Rapid climate change-related growth decline at the southern range edge of *fagus sylvatica*. *Global Change Biology*, **12**, 2163-2174
- Kachroo, A., Fu, D.Q., Havens, W., Navarre, D., Kachroo, P. & Ghabrial, S.A. (2008) An oleic acid-mediated pathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean. *Mol Plant Microbe Interact*, **21**, 564-575
- Kachroo, A., Lapchyk, L., Fukushige, H., Hildebrand, D., Klessig, D. & Kachroo, P. (2003a) Plastidial fatty acid signaling modulates salicylic acid- and jasmonic acid-mediated defense pathways in the *arabidopsis* *ssi2* mutant. *Plant Cell*, **15**, 2952-2965

- Kachroo, A., Shanklin, J., Whittle, E., Lapchyk, L., Hildebrand, D. & Kachroo, P. (2007) The arabidopsis stearyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. *Plant Mol Biol*, **63**, 257-271
- Kachroo, A., Venugopal, S.C., Lapchyk, L., Falcone, D., Hildebrand, D. & Kachroo, P. (2004) Oleic acid levels regulated by glycerolipid metabolism modulate defense gene expression in arabidopsis. *Proc Natl Acad Sci U S A*, **101**, 5152-5157
- Kachroo, P., Kachroo, A., Lapchyk, L., Hildebrand, D. & Klessig, D.F. (2003b) Restoration of defective cross talk in ssi2 mutants: Role of salicylic acid, jasmonic acid, and fatty acids in ssi2-mediated signaling. *Mol Plant Microbe Interact*, **16**, 1022-1029
- Kachroo, P., Shanklin, J., Shah, J., Whittle, E.J. & Klessig, D.F. (2001) A fatty acid desaturase modulates the activation of defense signaling pathways in plants. *Proc Natl Acad Sci U S A*, **98**, 9448-9453
- Kachroo, P., Venugopal, S.C., Navarre, D.A., Lapchyk, L. & Kachroo, A. (2005) Role of salicylic acid and fatty acid desaturation pathways in SSI2-mediated signaling. *Plant Physiol*, **139**, 1717-1735
- Kamal-Eldin, A. & Appelqvist, L. (1996a) The chemistry and antioxidant properties of tocopherols and tocotrienols. *Lipids*, **31**, 671-701
- Kamal-Eldin, A. & Appelqvist, L.A. (1996b) The chemistry and antioxidant properties of tocopherols and tocotrienols. *Lipids*, **31**, 671-701
- Kane, N. & Rieseberg, L. (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *helianthus annuus*. *Genetics*, **175**, 1823
- Kane, N. & Rieseberg, L. (2008) Genetics and evolution of weedy *helianthus annuus* populations: Adaptation of an agricultural weed. *Molecular Ecology*, **17**, 384-394
- Kang, B.T., Hauser, S. & Vanlauwe, B. (1995) Alley farming research on high status soils (ed. by A.O.O.a.a.L. Kang B.T.), pp. 25-39.

- Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J. & Eskin, E. (2008) Efficient control of population structure in model organism association mapping. *Genetics*, **178**, 1709-1723
- Kaplan, N.L., Hudson, R.R. & Langley, C.H. (1989) The "hitchhiking effect" revisited. *Genetics*, **123**, 887-899
- Keller, Y., Bouvier, F., D'harlingue, A. & Camara, B. (1998) Metabolic compartmentation of plastid prenyl lipid biosynthesis--evidence for the involvement of a multifunctional geranylgeranyl reductase. *Eur J Biochem*, **251**, 413-417
- Kelly, B., Bouvet, J.-M. & Picard, N. (2004a) Size class distribution and spatial pattern of *Vitellaria paradoxa* in relation to farmers' practices in mali. *Agroforestry Systems*, **60**, 3-11
- Kelly, B.A. (2005) Impact des pratiques humaines sur la dynamique des populations et sur la diversité génétique de *Vitellaria paradoxa* (karité) dans les systèmes agroforestiers au sud du Mali. In: *Population et Environnement*,. Université de Bamako, Bamako, Mali
- Kelly, B.A., Gourlet-Fleury, S. & Bouvet, J.M. (2007) Impact of agroforestry practices on the flowering phenology of *Vitellaria paradoxa* in parklands in southern mali. *Agroforestry Systems*, **71**, 67-75
- Kelly, B.A., Hardy, O. & Bouvet, J.M. (2004b) Temporal and spatial genetic structure in *Vitellaria paradoxa* (shea tree) in an agroforestry system in southern mali. *Mol Ecol*, **13**, 1231-1240
- Kelly, J.K. (1997) A test of neutrality based on interlocus associations. *Genetics*, **146**, 1197-1206
- Kettlewell, H.B.D. (1958) A survey of the frequencies of *Biston betularia* (l.) (lep.) and its melanic forms in great britain. *Heredity*, **12**, 51-72
- Khurana, E. & Singh, J.S. (2001) Ecology of tree seed and seedlings: Implications for tropical forest conservation and restoration. *Current Science*, **80**, 748-757

- Kim, Y. & Nielsen, R. (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics*, **167**, 1513-1524
- Kim, Y. & Stephan, W. (2000) Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics*, **155**, 1415-1427
- Kim, Y. & Stephan, W. (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160**, 765-777
- Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature*, **217**, 624-626
- Kimura, M. (1985) *The neutral theory of molecular evolution*. Cambridge Univ Pr.
- King, G., Maliepaard, C., Lynn, J., Alston, F., Durel, C., Evans, K., Griffon, B., Laurens, F., Manganaris, A. & Schrevers, E. (2000) Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*malus pumila* mill.). *TAG Theoretical and Applied Genetics*, **100**, 1074-1084
- Klein, B.P. & Kurilich, A.C. (2000) Processing effects on dietary antioxidants from plant foods. *Hortscience*, **35**, 580-584
- Knutzon, D.S., Scherer, D.E. & Schreckengost, W.E. (1991) Nucleotide sequence of a complementary DNA clone encoding stearyl-acyl carrier protein desaturase from castor bean, *ricinus communis*. *Plant Physiol*, **96**, 344-345
- Knutzon, D.S., Thompson, G.A., Radke, S.E., Johnson, W.B., Knauf, V.C. & Kridl, J.C. (1992) Modification of brassica seed oil by antisense expression of a stearyl-acyl carrier protein desaturase gene. *Proc Natl Acad Sci U S A*, **89**, 2624-2628
- Kotschy, T. (1865) *De plantis nilotico-aethiopicis knoblechterianis*. K. Akademie der Wissenschaften.
- Kruskal, J. (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, **7**, 48-50
- Krutovsky, K.V. & Neale, D.B. (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in douglas fir. *Genetics*, **171**, 2029-2041

- Ky, C.L., Barre, P., Lorieux, M., Trouslot, P., Akaffou, S., Louarn, J., Charrier, A., Hamon, S. & Noirot, M. (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theoretical and Applied Genetics*, **101**, 669-676
- Lam, H. (1939) On the system of the sapotaceae with some remarks on taxonomical methods. *Rec. Trav. bot. néerland*, **36**, 509-525
- Lamb, A. (1941) Different soils make different forests. *Farm and Forest*, **2**, 144-47
- Lamien, N., Sidibé, A. & Bayala, J. (1996) Use and commercialization of non-timber forest products in western Burkina Faso. *Domestication and commercialization of non-timber forest products in agroforestry systems, non-wood forest products* (ed. by T.A. Leakey Rrb, Melnyk M, Vantomme P), pp. 51-64. FAO, Rome, Italy.
- Langdale-Brown, I., Osmaston, H. & Wilson, J. (1964) *The vegetation of uganda and its bearing on land-use*. Govt. of Uganda.
- Langhelle, O. (2000) Sustainable development and social justice: Expanding the rawlsian framework of global justice. *Environmental Values*, **9**, 295-323
- Latta, R.G. (2004) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist*, **161**, 51-58
- Lavedrine, F., Ravel, A., Poupard, A. & Alary, J. (1997) Effect of geographic origin, variety and storage on tocopherol concentrations in walnuts by hplc. *Food Chemistry*, **58**, 135-140
- Le Corre, V. & Kremer, A. (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics*, **164**, 1205-1219
- Lely, H. (1925) *The useful trees of northern Nigeria*. Crown agents for the colonies.
- Lewicki, T. & Johnson, M. (1974) *West african food in the middle ages: According to arabic sources*. Cambridge University Press, Londres, GB.
- Librado, P. & Rozas, J. (2009) Dnasp v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452



- Liendo, R., Padilla, F.C. & Quintana, A. (1997) Characterization of cocoa butter extracted from criollo cultivars of *Theobroma cacao* L. *Food Research International*, **30**, 727-731
- Lightner, J., Wu, J. & Browse, J. (1994) A mutant of arabidopsis with increased levels of stearic acid. *Plant Physiol*, **106**, 1443-1451
- Linder, C.R. (2000) Adaptive evolution of seed oils in plants: Accounting for the biogeographic distribution of saturated and unsaturated fatty acids in seed oils. *American Naturalist*, **156**, 442-458
- Lindqvist, Y., Huang, W., Schneider, G. & Shanklin, J. (1996) Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *EMBO J*, **15**, 4081-4092
- Lioubimtseva, E., Simon, B., Faure, H., Faure-Denard, L. & Adams, J.M. (1998) Impacts of climatic change on carbon storage in the sahara-gobi desert belt since the last glacial maximum. *Global and Planetary Change*, **16-17**, 95-105
- Lloyd, A.H. (2005) Ecological histories from alaskan tree lines provide insight into future change. *Ecology*, **86**, 1687-1695
- Loehle, C. (1988) Tree life history strategies: The role of defenses. *Can. J. For. Res*, **18**, 209-222
- Lovett, P. & Haq, N. (2000a) Evidence for anthropic selection of the sheanut tree (*Vitellaria paradoxa*). *Agroforestry Systems*, **48**, 273-288
- Lovett, P.N. & Haq, N. (2000b) Diversity of the sheanut tree (*Vitellaria paradoxa* c.F. Gaertn.) in ghana. *Genetic Resources and Crop Evolution*, **47**, 293-304
- Lowe, A., Harris, D., Dormontt, E. & Dawson, I. (2010) Testing putative african tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Tropical Plant Biology*, **3**, 50-58
- Lynch, M. & Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics*, **154**, 459-473

- Maeda, H., Song, W., Sage, T.L. & Dellapenna, D. (2006) Tocopherols play a crucial role in low-temperature adaptation and phloem loading in arabidopsis. *Plant Cell*, **18**, 2710-2732
- Maley, J. (1989) Late quaternary climatic changes in the african rain forest : Forest refugia and the major role of sea surface temperature variations. *Paleoclimatology and paleometeorology : Modern and past patterns of global atmospheric transport* (ed. by S.M.E. Leinen M. (Ed.)), pp. p. 585-616. NATO ASI Series.Series C : Mathematical and Physical Sciences.
- Maley, J. (1991) The african rain forest vegetation and palaeoenvironments during late quaternary. *Climatic Change*, **19**, 79-98
- Maley, J. (1996) The african rain forest: Main characteristics of changes in vegetation and climate from the upper cretaceous to the quaternary. *Proceedings of the Royal Society of Edinburgh. Section B*, **104**, p. 31-73
- Mar-Heyming, R., Miyazaki, M., Weissglas-Volkov, D., Kolaitis, N.A., Sadaat, N., Plaisier, C., Pajukanta, P., Cantor, R.M., De Bruin, T.W., Ntambi, J.M. & Lusi, A.J. (2008) Association of stearoyl-coa desaturase 1 activity with familial combined hyperlipidemia. *Arterioscler Thromb Vasc Biol*, **28**, 1193-1199
- Maranz, S. & Wiesman, Z. (2003) Evidence for indigenous selection and distribution of the shea tree, *Vitellaria paradoxa*, and its potential significance to prevailing parkland savanna tree patterns in sub-saharan africa north of the equator. *Journal of Biogeography*, **30**, 1505-1516
- Maranz, S. & Wiesman, Z. (2004) Influence of climate on the tocopherol content of shea butter. *J Agric Food Chem*, **52**, 2934-2937
- Maranz, S., Wiesman, Z., Bisgaard, J. & Bianchi, G. (2004) Germplasm resources of *Vitellaria paradoxa* based on variations in fat composition across the species distribution range. *Agroforestry Systems*, **60**, 71-76
- Maranz, S., Wiesman, Z. & Garti, N. (2003) Phenolic constituents of shea (*Vitellaria paradoxa*) kernels. *J Agric Food Chem*, **51**, 6268-6273

- Marshall, P.S., Morris, S.R. & Threlfall, D.R. (1985) Biosynthesis of tocopherols: A re-examination of the biosynthesis and metabolism of 2-methyl-6-phytyl-1,4-benzoquinol. *Phytochemistry*, **24**, 1705-1711
- Marwede, V., Schierholt, A., Mollers, C. & Becker, H.C. (2004) Genotype x environment interactions and heritability of tocopherol contents in canola. *Crop Science*, **44**, 728-731
- Masters, E. (1992) Shea in uganda: A covol preliminary report. In. Cooperative Office for Voluntary Organizations of Uganda, Kampala, Uganda
- Matlin, A.J., Clark, F. & Smith, C.W.J. (2005) Understanding alternative splicing: Towards a cellular code. *Nat Rev Mol Cell Biol*, **6**, 386-398
- Maynard-Smith, J. & Haigh, J. (1974) The hitch-hiking effect of a favourable gene. *Genetics Research*, **23**, 23-35
- McDonald, J.H. & Kreitman, M. (1991) Adaptive protein evolution at the adh locus in drosophila. *Nature*, **351**, 652-654
- McKay, J.K. & Latta, R.G. (2002) Adaptive population divergence: Markers, qtl and traits. *Trends in Ecology & Evolution*, **17**, 285-291
- McVean, G. (2007) The structure of linkage disequilibrium around a selective sweep. *Genetics*, **175**, 1395-1406
- Meister, G. & Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded rna. *Nature*, **431**, 343-349
- Mène-Saffrané, L. & Dellapenna, D. (2010) Biosynthesis, regulation and functions of tocochromanols in plants. *Plant Physiology and Biochemistry*, **48**, 301-309
- Miège, J. (1954) Nombres chromosomiques et répartition géographique de quelques plantes tropicales et équatoriales. *Rev. Cytol. Biol. Vég.*, **15**, 312-348
- Moche, M., Shanklin, J., Ghoshal, A. & Lindqvist, Y. (2003) Azide and acetate complexes plus two iron-depleted crystal structures of the di-iron enzyme delta9 stearyl-acyl carrier

- protein desaturase. Implications for oxygen activation and catalytic intermediates. *J Biol Chem*, **278**, 25072-25080
- Moore, R.C. & Purugganan, M.D. (2005) The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology*, **8**, 122-128
- Mordelet, P., Abbadie, L. & Menaut, J. (1993) Effects of tree clumps on soil characteristics in a humid savanna of west Africa (Iamto, Cote d'Ivoire). *Plant and Soil*, **153**, 103-111
- Morley, R. & Richards, K. (1993) Gramineae cuticle: A key indicator of late cenozoic climatic change in the Niger delta. *Review of Palaeobotany and Palynology*, **77**, 119-127
- Morton, N.E., Zhang, W., Taillon-Miller, P., Ennis, S., Kwok, P.-Y. & Collins, A. (2001) The optimal measure of allelic association. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5217-5221
- Müller, F., Voccia, M., Bâ, A. & Bouvet, J. (2009) Genetic diversity and gene flow in a caribbean tree *Pterocarpus officinalis* jacq.: A study based on chloroplast and nuclear microsatellites. *Genetica*, **135**, 185-198
- Muller, J. (1981) Fossil pollen records of extant angiosperms. *The botanical review*, **47**, 1-142
- Nathan, R., Katul, G.G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W. & Levin, S.A. (2002) Mechanisms of long-distance dispersal of seeds by wind. *Nature*, **418**, 409-413
- Navascués, M. & Emerson, B.C. (2005) Chloroplast microsatellites: Measures of genetic diversity and the effect of homoplasmy. *Molecular Ecology*, **14**, 1333-1341
- Neale, D.B. & Savolainen, O. (2004) Association genetics of complex traits in conifers. *Trends in Plant Science*, **9**, 325-330
- Nei, M. (1969) Gene duplication and nucleotide substitution in evolution.
- Nei, M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321-3323
- Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.



- Neumann, K., Kahlheber, S. & Uebel, D. (1998) Remains of woody plants from saouga, a medieval west african village. *Vegetation History and Archaeobotany*, **7**, 57-77
- Nielsen, R. (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641-647
- Nielsen, R. (2005) Molecular signatures of natural selection. *Annu. Rev. Genet.*, **39**, 197-218
- Nordborg, M. & Weigel, D. (2008) Next-generation genetics in plants. *Nature*, **456**, 720-723
- Norris, S.R., Shen, X. & Della Penna, D. (1998) Complementation of the arabidopsis pds1 mutation with the gene encoding p-hydroxyphenylpyruvate dioxygenase. *Plant Physiol.*, **117**, 1317-1323
- Ogundiwin, E., Peace, C., Gradziel, T., Parfitt, D., Bliss, F. & Crisosto, C. (2009) A fruit quality gene map of *Prunus*. *BMC Genomics*, **10**, 587
- Ohno, S. (1967) *Sex chromosomes and sex-linked genes*. Berlin, Germany: Springer.
- Ohno, S. (1970) *Evolution by gene duplication*. New York, NY, USA: Springer.
- Okada, K., Saito, T., Nakagawa, T., Kawamukai, M. & Kamiya, Y. (2000) Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in arabidopsis. *Plant Physiol*, **122**, 1045-1056
- Okubo, A. & Levin, S.A. (1989) A theoretical framework for data analysis of wind dispersal of seeds and pollen. *Ecology*, **70**, 329-338
- Okullo, J., Omujal, F., Agea, J., Vuzi, P., Namutebi, A., Okello, J. & Nyanzi, S. (2010) Physico-chemical characteristics of shea butter (*Vitellaria paradoxa* cf gaertn.) oil from the shea district of uganda. *African Journal of Food, Agriculture, Nutrition and Development*, **10**
- Oldeman, L. & Van Engelen, V. (1993) A world soils and terrain digital database (soter)--an improved assessment of land resources. *Geoderma*, **60**, 309-325
- Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. & De Silva, H.N. (2007) *Association mapping in plants*. Springer Science + Business Media, LLC.

- Ordóñez, J.L., Retana, J. & Espelta, J.M. (2005) Effects of tree size, crown damage, and tree location on post-fire survival and cone production of *Pinus nigra* trees. *Forest Ecology and Management*, **206**, 109-117
- Orr, H. (2005) The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, **6**, 119-127
- Osei-Amaning, E. (1996) Management of *Vitellaria paradoxa* in guinea savanna rangelands in ghana. In, p. 199 University of Wales, Bangor, GB
- Otto-Bliesner, B.L., Marshall, S.J., Overpeck, J.T., Miller, G.H., Hu, A. & Members, C.L.I.P. (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science*, **311**, 1751-1753
- Palm, C. (1995) Contribution of agroforestry trees to nutrient requirements of intercropped plants. *Agroforestry Systems*, **30**, 105-124
- Park, M. & Rennell, J. (1799) Travels in the interior districts of Africa: Performed under the direction and patronage of the African Association, in the years 1795, 1796, and 1797. *Travels into the interior of africa* (ed. by S. J), pp. 1-264. Eland, Londre, GB.
- Parks, D.H., Porter, M., Churcher, S., Wang, S., Blouin, C., Whalley, J., Brooks, S. & Beiko, R.G. (2009) Gengis: A geospatial information system for genomic data. *Genome Res*, **19**, 1896-1904
- Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637-669
- Parry, M. (2007) *Climate change 2007: Impacts, adaptation and vulnerability: Contribution of working group ii to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge Univ Pr.
- Pasquini, C. (2003) Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, **14**, 198-219
- Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225-245

- Pennington, T. & Krukoff, B. (1991) *The genera of sapotaceae*. Royal Botanic Gardens, Kew, New York.
- Petit, R., My, J., Hampe, A. & Cheddadi, R. (2005) Climate changes and tree phylogeography in the mediterranean. *Taxon*, **54**, 877-885
- Petit, R.J., Aguinagalde, I., De Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Muller-Starck, G., Demesure-Musch, B., Palme, A., Martin, J.P., Rendell, S. & Vendramin, G.G. (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563-1565
- Petit, R.J. & Hampe, A. (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics*, **37**, 187-214
- Petit, R.J., Hu, F.S. & Dick, C.W. (2008) Forests of the past: A window to future changes. *Science*, **320**, 1450-1452
- Petit, R.J., Mousadik, A.E. & Pons, O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844-855
- Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducousso, A. & Kremer, A. (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc Natl Acad Sci U S A*, **94**, 9996-10001
- Philippe, R., Courtois, B., McNally, K., Mournet, P., El-Malki, R., Le Paslier, M., Fabre, D., Billot, C., Brunel, D., Glaszmann, J.-C. & This, D. (2010) Structure, allelic diversity and selection of asr genes, candidate for drought tolerance, in *Oryza sativa* L. And wild relatives. *TAG Theoretical and Applied Genetics*, **121**, 769-787
- Phillips, S., Dudík, M. & Schapire, R. (2004) A maximum entropy approach to species distribution modeling. In, p. 83. ACM
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231-259
- Pilz, G.E. (1980) Chromosome number reports lxvii. *Taxon*, **29**, 352-353

- Pischikova, E. (2009) Early kushite tombs of south asaf. *British Museum Studies in Ancient Egypt and Sudan*, **12**, 11-30
- Pisias, N.G. & Moore Jr, T.C. (1981) The evolution of pleistocene climate: A time series approach. *Earth and Planetary Science Letters*, **52**, 450-458
- Pleite, R., Martinez-Force, E. & Garces, R. (2006) Increase of the stearic acid content in high-oleic sunflower (*Helianthus annuus*) seeds. *J Agric Food Chem*, **54**, 9383-9388
- Pock Tsy, J.-M.L., Lumaret, R., Mayne, D., Vall, A.O.M., Abutaba, Y.I.M., Sagna, M., Raoseta, S.N.O.R. & Danthu, P. (2009) Chloroplast DNA phylogeography suggests a west african centre of origin for the baobab, *Adansonia digitata* L. (bombacoideae, malvaceae). *Molecular Ecology*, **18**, 1707-1715
- Pons, O. & Petit, R.J. (1995) Estimation, variance and optimal sampling of gene diversity. *TAG Theoretical and Applied Genetics*, **90**, 462-470
- Pons, O. & Petit, R.J. (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237-1245
- Porfirova, S., Bergmüller, E., Tropf, S., Lemke, R. & Dörmann, P. (2002) Isolation of an arabidopsis mutant lacking vitamin e and identification of a cyclase essential for all tocopherol biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 12495-12500
- Portnoy, S. & Willson, M.F. (1993) Seed dispersal curves: Behavior of the tail of the distribution. *Evolutionary Ecology*, **7**, 25-44
- Poumot, C. (1989) Palynological evidence for eustatic events in the tropical neogene. *Bulletin des centres de recherches exploration-production ELF Aquitaine, Boussens, 1989*, 13 (2), 437-453
- Primomo, V.S., Falk, D.E., Ablett, G.R., Tanner, J.W. & Rajcan, I. (2002) Genotype x environment interactions, stability, and agronomic performance of soybean with altered fatty acid profiles. *Crop Sci*, **42**, 37-44

- Pritchard, H.W., Daws, M.I., Fletcher, B.J., Gamene, C.S., Msanga, H.P. & Omondi, W. (2004) Ecological correlates of seed desiccation tolerance in tropical african dryland trees. *American Journal of Botany*, **91**, 863-870
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959
- Prokarité [Http://Www.Prokarite.Org/](http://Www.Prokarite.Org/)
- Przeworski, M. (2002) The signature of positive selection at randomly chosen loci. *Genetics*, **160**, 1179-1189
- Quattrocchio, F., Wing, J., Van Der Woude, K., Souer, E., De Vetten, N., Mol, J. & Koes, R. (1999) Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. *Plant Cell*, **11**, 1433-1444
- Rafalski, J.A. (2002) Novel genetic mapping tools in plants: Snps and Id-based approaches. *Plant Science*, **162**, 329-333
- Raintree, J. (1987) The state of the art of agroforestry diagnosis and design. *Agroforestry Systems*, **5**, 219-250
- Ramos, A.C.S., De Lemos-Filho, J.P. & Lovato, M.B. (2009) Phylogeographical structure of the neotropical forest tree *Hymenaea courbaril* (leguminosae: Caesalpinioideae) and its relationship with the vicariant *hymenaea stigonocarpa* from cerrado. *Journal of Heredity*, **100**, 206-216
- Ramos, A.C.S., Lemos-Filho, J.P., Ribeiro, R.A., Santos, F.R. & Lovato, M.B. (2007) Phylogeography of the tree *Hymenaea stigonocarpa* (fabaceae: Caesalpinioideae) and the influence of quaternary climate changes in the brazilian cerrado. *Annals of Botany*, **100**, 1219-1228
- Rebetzke, G., Burton, J., Carter, T. & Wilson, R. (1998) Genetic variation for modifiers controlling reduced saturated fatty acid content in soybean. *Crop Science*, **38**, 303-308

- Rees, M. & Long, M.J. (1992) Germination biology and the ecology of annual plants. *American Naturalist*, **139**, 484-508
- Reich, D.E. & Goldstein, D.B. (1998) Genetic evidence for a paleolithic human population expansion in africa. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 8119-8123
- Rhone, B., Remoue, C., Galic, N., Goldringer, I. & Bonnin, I. (2008) Insight into the genetic bases of climatic adaptation in experimentally evolving wheat populations. *Molecular Ecology*, **17**, 930-943
- Rieseberg, L.H. & Blackman, B.K. (2010) Speciation genes in plants. *Annals of Botany*, **106**, 439-455
- Ritland, K. & Travis, S. (2004) Inferences involving individual coefficients of relatedness and inbreeding in natural populations of *Abies*. *Forest Ecology and Management*, **197**, 171-180
- Roark, B. & Quisenberry, J.E. (1977) Environmental and genetic components of stomatal behavior in two genotypes of upland cotton. *Plant Physiol.*, **59**, 354-356
- Rocheleau, D., Weber, F. & Field-Juma, A. (1988) Agroforestry in dryland Africa.
- Rodin, S. & Riggs, A. (2003) Epigenetic silencing may aid evolution by gene duplication. *Journal of Molecular Evolution*, **56**, 718-729
- Rogers, A.R. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552-569
- Rowe, N. & Speck, T. (2005) Plant growth forms: An ecological and evolutionary perspective. *New Phytol*, **166**, 61-72
- Ruyssen, B. (1957) Le karité au soudan. *Agronomie Tropicale*, **12**, 143–171, 279-306, 415-440
- Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R. & Lander, E.S. (2002) Detecting recent

- positive selection in the human genome from haplotype structure. *Nature*, **419**, 832-837
- Sallé, G., Boussim, J., Raynal-Roques, A. & Brunck, F. (1991) Le karité: État de nos connaissances et perspectives de recherche. *Physiologie des arbres et arbustes en zones arides et semi-arides* (ed. by E.D. A. Riedacker, C. Pafadnam, H. Joly, G. Bory), pp. 427-439. John Libbey, Paris.
- Sankaran, M., Hanan, N.P., Scholes, R.J., Ratnam, J., Augustine, D.J., Cade, B.S., Gignoux, J., Higgins, S.I., Le Roux, X., Ludwig, F., Ardo, J., Banyikwa, F., Bronn, A., Bucini, G., Caylor, K.K., Coughenour, M.B., Diouf, A., Ekaya, W., Feral, C.J., February, E.C., Frost, P.G., Hiernaux, P., Hrabar, H., Metzger, K.L., Prins, H.H., Ringrose, S., Sea, W., Tews, J., Worden, J. & Zambatis, N. (2005) Determinants of woody cover in African savannas. *Nature*, **438**, 846-849
- Sankaran, M., Ratnam, J. & Hanan, N.P. (2004) Tree–grass coexistence in savannas revisited – insights from an examination of assumptions and mechanisms invoked in existing models. *Ecology Letters*, **7**, 480-490
- Sanou, H., Lovett, P.N. & Bouvet, J.M. (2005) Comparison of quantitative and molecular variation in agroforestry populations of the shea tree (*Vitellaria paradoxa* C.F. Gaertn.) in Mali. *Mol Ecol*, **14**, 2601-2610
- Sanou, H., Picard, N., Lovett, P.N., Dembele, M., Korbo, A., Diarisso, D. & Bouvet, J.M. (2006) Phenotypic variation of agromorphological traits of the shea tree, *Vitellaria paradoxa* cf gaertn., in mali. *Genetic Resources and Crop Evolution*, **53**, 145-161
- Sarmiento, G. (1984) *The ecology of neotropical savannas*. Harvard Univ Pr, Cambridge.
- Sarmiento, G. (1992) A conceptual model relating environmental factors and vegetation formations in the lowlands of tropical south america. *Nature and dynamics of forest-savanna boundaries* (ed. by P.a.P. Furley, J.; Ratter, Ja). Chapman and Hall, London, GB.

- Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M. & Dellapenna, D. (2004) Vitamin e is essential for seed longevity and for preventing lipid peroxidation during germination. *Plant Cell*, **16**, 1419-1432
- Savidge, B., Weiss, J.D., Wong, Y.H., Lassner, M.W., Mitsky, T.A., Shewmaker, C.K., Post-Beittenmiller, D. & Valentin, H.E. (2002) Isolation and characterization of homogentisate phytyltransferase genes from *synechocystis* sp. Pcc 6803 and *arabidopsis*. *Plant Physiol*, **129**, 321-332
- Savolainen, O. & Pyhäjärvi, T. (2007) Genomic diversity in forest trees. *Current Opinion in Plant Biology*, **10**, 162-167
- Sawada, H., Kuykendall, L.D. & Young, J.M. (2003) Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. *The Journal of General and Applied Microbiology*, **49**, 155-179
- Schäfer, K. (1998) Accelerated solvent extraction of lipids for determining the fatty acid composition of biological material. *Analytica Chimica Acta*, **358**, 69-77
- Schierholt, A. & Becker, H.C. (2001) Environmental variability and heritability of high oleic acid content in winter oilseed rape. *Plant Breeding*, **120**, 63-66
- Schlotterer, C. (2003) Hitchhiking mapping--functional genomics from the population genetics perspective. *Trends Genet*, **19**, 32-38
- Schneider, G., Lindqvist, Y., Shanklin, J. & Somerville, C. (1992) Preliminary crystallographic data for stearoyl-acyl carrier protein desaturase from castor seed. *J Mol Biol*, **225**, 561-564
- Schneider, S. & Excoffier, L. (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics*, **152**, 1079-1089
- Scholes, R. & Walker, B. (1993) *An African savanna: Synthesis of the nylsvley study*. Cambridge University Press, Cambridge.

- Scholes, R.J. & Archer, S.R. (1997) Tree-grass interactions in savannas. *Annual Review of Ecology and Systematics*, **28**, 517-544
- Schreckenberg, K. (1996) Forests, fields and markets a study of indigenous tree products in the woody savannas of the bassila region, benin. In, p. 326. University of London, Londres, GB
- Schroth, G. (1998) A review of belowground interactions in agroforestry, focussing on mechanisms and management options. *Agroforestry Systems*, **43**, 5-34
- Sebillotte, M. (1985) La jachère : Éléments pour une théorie. *Dynamique des systèmes agraires : À travers champs agronomes et géographes*, pp. 175-229. ORSTOM, Paris.
- Serpantié, G. (1996) Rôles des jachères dans la production arborée non ligneuse en savane soudanienne : Cas du karité dans l'ouest du burkina faso. In: *La Jachère, Lieu de Production : Atelier* (ed. C. Floret), pp. 55-57 multigr. ORSTOM, Dakar
- Shah, J., Kachroo, P., Nandi, A. & Klessig, D.F. (2001) A recessive mutation in the arabidopsis *SSI2* gene confers sa- and npr1-independent expression of *PR* genes and resistance against bacterial and oomycete pathogens. *Plant J*, **25**, 563-574
- Shanklin, J., Mullins, C. & Somerville, C. (1991) Sequence of a complementary DNA from *cucumis sativus* L. Encoding the stearyl-acyl-carrier protein desaturase. *Plant Physiol*, **97**, 467-468
- Shanklin, J. & Somerville, C. (1991) Stearyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. *Proc Natl Acad Sci U S A*, **88**, 2510-2514
- Sharma, T. (1976) The pre-historic background of shifting cultivation in ne india. *North East India Council for Social Sciences Research, Shillong, India*, 1-5
- Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E. & Small, R.L. (2005) The tortoise and the hare ii: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142-166

- Shintani, D. & Dellapenna, D. (1998) Elevating the vitamin e content of plants through metabolic engineering. *Science*, **282**, 2098-2100
- Shintani, D.K., Cheng, Z. & Dellapenna, D. (2002) The role of 2-methyl-6-phytylbenzoquinone methyltransferase in determining tocopherol composition in *synechocystis* sp. Pcc6803. *FEBS Lett*, **511**, 1-5
- Silvertown, J. (1989) The paradox of seed size and adaptation. *Trends in Ecology & Evolution*, **4**, 24-26
- Siol, M., Wright, S.I. & Barrett, S.C.H. (2010) The population genomics of plant adaptation. *New Phytologist*, **188**, 313-332
- Skarpe, C. (1992) Dynamics of savanna ecosystems. *Journal of Vegetation Science*, **3**, 293-300
- Smith, W. (1933) The coffin of prince min-khaf. *The Journal of Egyptian Archaeology*, **19**, 150-159
- Soll, J. (1987) Alpha-tocopherol and plastoquinone synthesis in chloroplast membranes. *Methods in enzymology* (ed. by R.D. Lester Packer), pp. 383-392. Academic Press.
- Soll, J., Kemmerling, M. & Schultz, G. (1980) Tocopherol and plastoquinone synthesis in spinach chloroplasts subfractions. *Arch Biochem Biophys*, **204**, 544-550
- Solomon, S. (2007) *Climate change 2007: The physical science basis: Contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge Univ Pr.
- Somerville, C., Browse, J., Jaworski, J.G., Ohlrogge, J.B. & (Buchanan, B., Jones, R., and Gruissem, W., Eds) (2000) *Chapter 10 lipids in biochemistry and molecular biology of plants* American Society of Plant Physiologists, Rockville, MD, USA.
- Sorensen, F. (1999) Relationship between self-fertility, allocation of growth, and inbreeding depression in three coniferous species. *Evolution*, **53**, 417-425
- Soria-Auza, R.W., Kessler, M., Bach, K., Barajas-Barbosa, P.M., Lehnert, M., Herzog, S.K. & Böhner, J. (2010) Impact of the quality of climate models for modelling species



- occurrences in countries with poor climatic documentation: A case study from bolivia. *Ecological Modelling*, **221**, 1221-1229
- Spring, J. (2002) Genome duplication strikes back. *Nature Genetics*, **31**, 128-129
- Stephan, W., Wiehe, T.H.E. & Lenz, M.W. (1992) The effect of strongly selected substitutions on neutral polymorphism - analytical results based on diffusion-theory. *Theoretical Population Biology*, **41**, 237-254
- Stinchcombe, J.R. & Hoekstra, H.E. (2008) Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity*, **100**, 158-170
- Stocker, A., Fretz, H., Frick, H., Ruttimann, A. & Woggon, W.D. (1996) The substrate specificity of tocopherol cyclase. *Bioorg Med Chem*, **4**, 1129-1134
- Storz, J.F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671-688
- Swenson, U. & Anderberg, A.A. (2005) Phylogeny, character evolution, and classification of sapotaceae (Ericales). *Cladistics*, **21**, 101-130
- Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105-1109
- Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437-460
- Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595
- Takhtajan, A. (1980) Outline of the classification of flowering plants (magnoliophyta). *The botanical review*, **46**, 225-359
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) Mega4: Molecular evolutionary genetics analysis (mega) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599

- Tarasov, P.E., Volkova, V.S., Webb, T., Guiot, J., Andreev, A.A., Bezusko, L.G., Bezusko, T.V., Bykova, G.V., Dorofeyuk, N.I., Kvavadze, E.V., Osipova, I.M., Panova, N.K. & Sevastyanov, D.V. (2000) Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern eurasia. *Journal of Biogeography*, **27**, 609-620
- Teklehaimanot, Z. (2004) Exploiting the potential of indigenous agroforestry trees: *Parkia biglobosa* and *Vitellaria paradoxa* in sub-saharan africa. *Agroforestry Systems*, **61**, 207-220
- Thompson, G.A., Scherer, D.E., Foxall-Van Aken, S., Kenny, J.W., Young, H.L., Shintani, D.K., Kridl, J.C. & Knauf, V.C. (1991) Primary structures of the precursor and mature forms of stearyl-acyl carrier protein desaturase from safflower embryos and requirement of ferredoxin for enzyme activity. *Proc Natl Acad Sci U S A*, **88**, 2578-2582
- Tong, L., Shu-Ming, P., Wu-Yuan, D., Dan-Wei, M., Ying, X., Meng, X. & Fang, C. (2006) Characterization of a new stearyl-acyl carrier protein desaturase gene from *jatropha curcas*. *Biotechnol Lett*, **28**, 657-662
- Traoré, K.G., Francis, Oliver, Robert; Gigou, Jacques (2004) Litter production and soil fertility in a *Vitellaria paradoxa* parkland in a catena in southern mali. *Arid Land Research and Management*, **18**, 359-368
- Tripiana, V., Bourgeois, M., Verhaegen, D., Vigneron, P. & Bouvet, J.-M. (2007) Combining microsatellites, growth, and adaptive traits for managing in situ genetic resources of *Eucalyptus urophylla*. *Canadian Journal of Forest Research*, **37**, 773-785
- Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroove, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang,

- W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van De Peer, Y. & Rokhsar, D. (2006) The genome of black cottonwood, *Populus trichocarpa* (torr. & gray). *Science*, **313**, 1596-1604
- Van Langevelde, F., Van De Vijver, C.a.D.M., Kumar, L., Van De Koppel, J., De Ridder, N., Van Andel, J., Skidmore, A.K., Hearne, J.W., Stroosnijder, L., Bond, W.J., Prins, H.H.T. & Rietkerk, M. (2003) Effects of fire and herbivory on the stability of savanna ecosystems. *Ecology*, **84**, 337-350
- Vanin, E.F. (1985) Processed pseudogenes: Characteristics and evolution. *Annual Review of Genetics*, **19**, 253-272
- Varshney, R.K., Nayak, S.N., May, G.D. & Jackson, S.A. (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology*, **27**, 522-530
- Velasco, L., Fernandez-Martinez, J.M., Garcia-Ruiz, R. & Dominguez, J. (2002) Genetic and environmental variation for tocopherol content and composition in sunflower commercial hybrids. *Journal of Agricultural Science*, **139**, 425-429
- Velioglu, Y.S., Mazza, G., Gao, L. & Oomah, B.D. (1998) Antioxidant activity and total phenolics in selected fruits, vegetables, and grain products. *Journal of Agricultural and Food Chemistry*, **46**, 4113-4117
- Vendramin, G.G., Fady, B., Gonzalez-Martinez, S.C., Hu, F.S., Scotti, I., Sebastiani, F., Soto, A. & Petit, R.J. (2008) Genetically depauperate but widespread: The case of an emblematic mediterranean pine. *Evolution*, **62**, 680-688

- Venegas-Caleron, M., Youssar, L., Salas, J.J., Garces, R. & Martinez-Force, E. (2009) Effect of the ferredoxin electron donor on sunflower (*Helianthus annuus*) desaturases. *Plant Physiol Biochem*, **47**, 657-662
- Venkatesh, T., Karunanandaa, B., Free, D., Rottnek, J., Baszis, S. & Valentin, H. (2006) Identification and characterization of an arabidopsis homogentisate phytyltransferase paralog. *Planta*, **223**, 1134-1144
- Verger, R. & Rivière, C. (1987) Les enzymes lipolytiques: Une étude cinétique *Rev. Fr. Corps Gras.* , **1**, pp. 7-13.
- Vertucci, C.W. & Leopold, A.C. (1987) Oxidative processes in soybean and pea seeds: Effect of light, temperature, and water content. *Plant Physiol*, **84**, 1038-1043
- Vigouroux, Y., Jaqueth, J.S., Matsuoka, Y., Smith, O.S., Beavis, W.D., Smith, J.S.C. & Doebley, J. (2002) Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution*, **19**, 1251-1260
- Vivien, J. (1990) Fruitiers sauvages du cameroun. *Fruits*, **45**, 149-160
- Vollhardt, K. & Schore, N. (2004) *Traité de chimie organique*. De Boeck Université.
- Vuillet, J. & Africa, F. (1911) *Le karité et ses produits*. É. Larose, Paris.
- Wachowiak, W., Balk, P. & Savolainen, O. (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in scots pine (*Pinus sylvestris* L.). *Tree Genetics & Genomes*, **5**, 117-132
- Walker, B.H. & Noy-Meir, I. (1982) Aspects of the stability and resilience of savanna ecosystems. *Ecology of tropical savannas* (ed. by B.J.H.E.B.H.W. (Eds)). Ecological Studies, Springer-Verlag, Berlin.
- Ward, J.K., Harris, J.M., Cerling, T.E., Wiedenhoef, A., Lott, M.J., Dearing, M.D., Coltrain, J.B. & Ehleringer, J.R. (2005) Carbon starvation in glacial trees recovered from the la brea tar pits, southern california. *Proc Natl Acad Sci U S A*, **102**, 690-694
- Wardle, D.A., Walker, L.R. & Bardgett, R.D. (2004) Ecosystem properties and forest decline in contrasting long-term chronosequences. *Science*, **305**, 509-513

- Waterman, P. & Mahmoud, E. (1991) Chemical taxonomy of the Sapotaceae: Patterns in the distribution of some simple phenolic compounds. *The genera of Sapotaceae* (ed. by T. Pennington), pp. 51-74. Royal Botanic Gardens Kew & New York Botanical Garden, New York.
- Watterson, G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*, **7**, 256-276
- Weising, K. & Gardner, R.C. (1999) A set of conserved pcr primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, **42**, 9-19
- Weiss, K.M. & Clark, A.G. (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends in Genetics*, **18**, 19-24
- Westoby, M., Jurado, E. & Leishman, M. (1992) Comparative evolutionary ecology of seed size. *Trends in Ecology & Evolution*, **7**, 368-372
- White, F. (1983) The vegetation of Africa: A descriptive memoir to accompany the unesco/aetfat/unso vegetation map of africa. *Natural resources research*, **20**, 1-136
- White, R., Murray, S., Rohweder, M. & Institute, W.R. (2000) *Pilot analysis of global ecosystems: Grassland ecosystems*. World Resources Institute.
- Whittle, E., Cahoon, E.B., Subrahmanyam, S. & Shanklin, J. (2005) A multifunctional acyl-acyl carrier protein desaturase from *hedera helix* L. (english ivy) can synthesize 16-and 18-carbon monoene and diene products. *Journal of Biological Chemistry*, **280**, 28169-28176
- Wickens, G. (1976) Speculations on long distance dispersal and the flora of jebel marra, sudan republic. *Kew Bulletin*, **31**, 105-150
- Wicker, F. (1998) The road to punt. *Geographical Journal*, **164**, 155-167
- Willyard, A., Syring, J., Gernandt, D.S., Liston, A. & Cronn, R. (2007) Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for. *Molecular Biology and Evolution*, **24**, 90-101

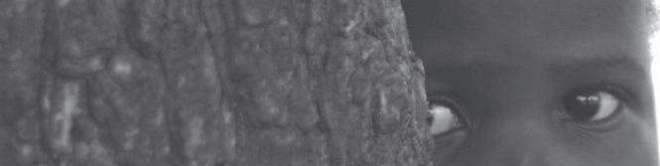


- Wolfe, K.H. (2001) Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics*, **2**, 333-341
- Wolfe, K.H., Li, W.H. & Sharp, P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear dnas. *Proc Natl Acad Sci U S A*, **84**, 9054-9058
- Woodward, F.I., Lomas, M.R. & Kelly, C.K. (2004) Global climate and the distribution of plant biomes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 1465-1476
- Xia, H., Camus-Kulandaivelu, L., Stephan, W., Tellier, A. & Zhang, Z.W. (2010) Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. *Molecular Ecology*, **19**, 4144-4154
- Yamamoto, E., Takashi, T., Morinaka, Y., Lin, S.Y., Wu, J.Z., Matsumoto, T., Kitano, H., Matsuoka, M. & Ashikari, M. (2010) Gain of deleterious function causes an autoimmune response and bateson-dobzhansky-muller incompatibility in rice. *Molecular Genetics and Genomics*, **283**, 305-315
- Yang, Z.H. & Bielawski, J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, **15**, 496-503
- Young, M. & Solbrig, O. (1993) *The world's savannas: Economic driving forces, ecological constraints and policy options for sustainable land use*. UNESCO/Parthenon Publ. Group.
- Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S. & Buckler, E.S. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*, **38**, 203-208
- Zhang, J.Z. (2003) Evolution by gene duplication: An update. *Trends in Ecology & Evolution*, **18**, 292-298
- Zhang, L.Q., Gaut, B.S. & Vision, T.J. (2001) Gene duplication and evolution. *Science*, **293**, U1-U2



- Zhang, R., Zheng, Y., Wu, Y., Wang, S. & Chen, F. (2006) A simple and efficient method for preparation of plant RNAs. *Yi chuan= Hereditas/Zhongguo yi chuan xue hui bian ji*, **28**, 583
- Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M. & Buckler, E.S. (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat Genet*, **42**, 355-360
- Zhao, K.Y., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C.L., Toomajian, C., Zheng, H.G., Dean, C., Marjoram, P. & Nordborg, M. (2007) An arabidopsis example of association mapping in structured samples. *Plos Genetics*, **3**, 4
- Zhen, Y. & Ungerer, M. (2008) Relaxed selection on the *cbf/dreb1* regulatory genes and reduced freezing tolerance in the southern range of *Arabidopsis thaliana*. *Molecular Biology and Evolution*, **25**, 2547
- Zobel, B. & Talbert, J. (1984) *Applied forest tree improvement*. John Wiley & Sons.





Annexes

ANNEXE II.1. PROTOCOLE DE PCR TYPE

Mix PCR par puits	mix /puits (µl)
tampon 10X (1X final = 50mM KCl + 10mM tris-HCl pH8,3)	2
MgCl ₂ 50mM (1,5mM final)	0,6
dNTPs 5mM chaque (0,2mM final)	0,8
amorce Upper+ queue M13, 10 µM (0,18 µM final)	0,36
amorce Lower 10 µM (0,30 final)	0,6
oligo M13 marqué 1µM (0,15 µM final)	3
eau	6,4
Polymérase Taq 5 U/µl (0,06 U/µl final)	0,24
Total	14 µl
ADN matrice 5 ng/µl (1,5 ng/µl final)	6

- Les plaques PCR sont recouvertes par un film spécial contenant de la gélose pour éviter l'évaporation.
- Machine PCR: MasterCycler Eppendorf® 384 puits
- 4 Programmes selon le locus en fonction de la température FA50/56/58/64:

1	94 °C	4 min		Dénaturation initiale	
2	94 °C	45 sec		Dénaturation	
3	55 °C	1 min	Moins 0.5 °C/cycle	Hybridation	10 cycles
4	72 °C	1 min 15 sec		Élongation	
5	94 °C	45 sec		Dénaturation	
6	50 °C	1 min		Hybridation	30 cycles
7	72 °C	1 min 15 sec		Élongation	
8	72 °C	5 min		Élongation Finale	

Contrôle:

- Couler un grand gel d'agarose à 2% (6 g d'agarose + 300 ml de TAE)
- Ajouter 2,5 µl de bleu stop à chaque puits d'amplification
- Déposer 18 µl par puits, faire migrer à 120 volts environ 4h
- Colorer le gel 10 mn au BET, prendre une photo

ANNEXE II.2. PROTOCOLE DE GÉNOTYPAGE DES MARQUEURS

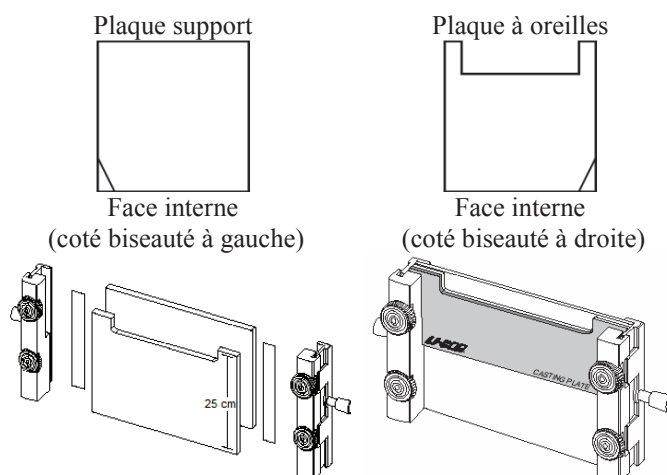
MICROSATELLITES

Préparation des plaques pour couler un gel de séquenceur

Nettoyer la face interne des plaques à l'éthanol à 90°. Mettre une goutte d'eau sur les « spacers » de 0.22 mm. Disposer les « spacers » sur la plaque support. Superposer les deux plaques (les deux encoches biseautées doivent être sur la gauche). Mettre la "casting plate" sur la plaque à oreilles. Serrer de part et d'autre des plaques avec les serre-joints à vis.

Coulage du gel

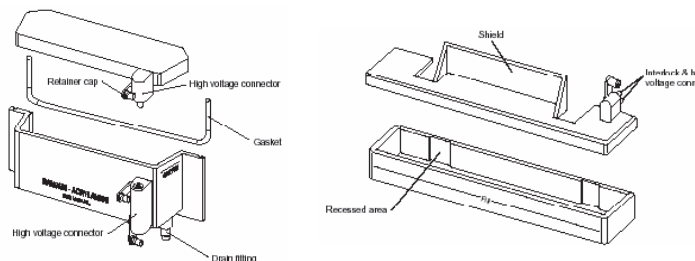
- Mettre 20 ml d'acrylamide Long Ranger (LR) à 6,5% dans un bécher en le faisant couler le long de la paroi, afin de ne pas entraîner la formation de bulles.
- Ajouter 25 µl de TEMED et 175 µl d'APS à 10 %.
- Agiter la solution en faisant lentement tourner le becher.
- Incliner les plaques.
- Couler le gel.
- Remettre les plaques à l'horizontale.
- Placer la « surface former » au centre de la plaque dans l'encoche.
- Resserrer les vis des serre-joints.
- Enlever le surplus d'acrylamide autour du « surface former » avec du papier sopalin.
- Laisser polymériser une heure au moins.



Montage du gel

Enlever les serre-joints à vis. Enlever la "casting plate". Enlever le "surface former". Nettoyer l'acrylamide à l'endroit du "surface former". Monter la cuve supérieure à la place de la "casting plate". Remettre les serre-joints. Nettoyer les faces extérieures des plaques à l'eau puis à l'éthanol à 90° (il ne doit pas rester de traces d'acrylamide, en particulier en face de la fenêtre de lecture). Mettre la cuve inférieure dans le LICOR. Monter le gel dans le LICOR. Préparer 1 L de TBE 1X. Remplir la cuve supérieure avec du TBE 1X jusqu'au trait. Nettoyer la zone de dépôt avec une seringue et un cône spatulé pour enlever les filaments d'acrylamide présents.

Remplir les cuves inférieures et supérieures avec du TBE 1X. Disposer les couvercles adaptés sur les cuves inférieure et supérieure. Brancher l'électrode du haut. Fermer le capot du LICOR.



Dépôt

Enlever l'électrode. Ôter le couvercle de la cuve supérieure. Nettoyer la surface du gel avec une seringue pour qu'il n'y ait pas de morceau d'acrylamide. Mettre en place le peigne papier (96 dents) chargé en produit PCR sans trop l'enfoncer pour ne pas déformer la surface du gel. Remettre le couvercle de la cuve supérieure. Brancher l'électrode. Cliquer sur « Collect image ».

ANNEXE II.3. DIVA-GIS POPULATIONS

Populations utilisées pour la modélisation des niches écologiques (ENM) de *Vitellaria paradoxa* par le logiciel DIVA-GIS (Hijmans *et al.*, 2001).

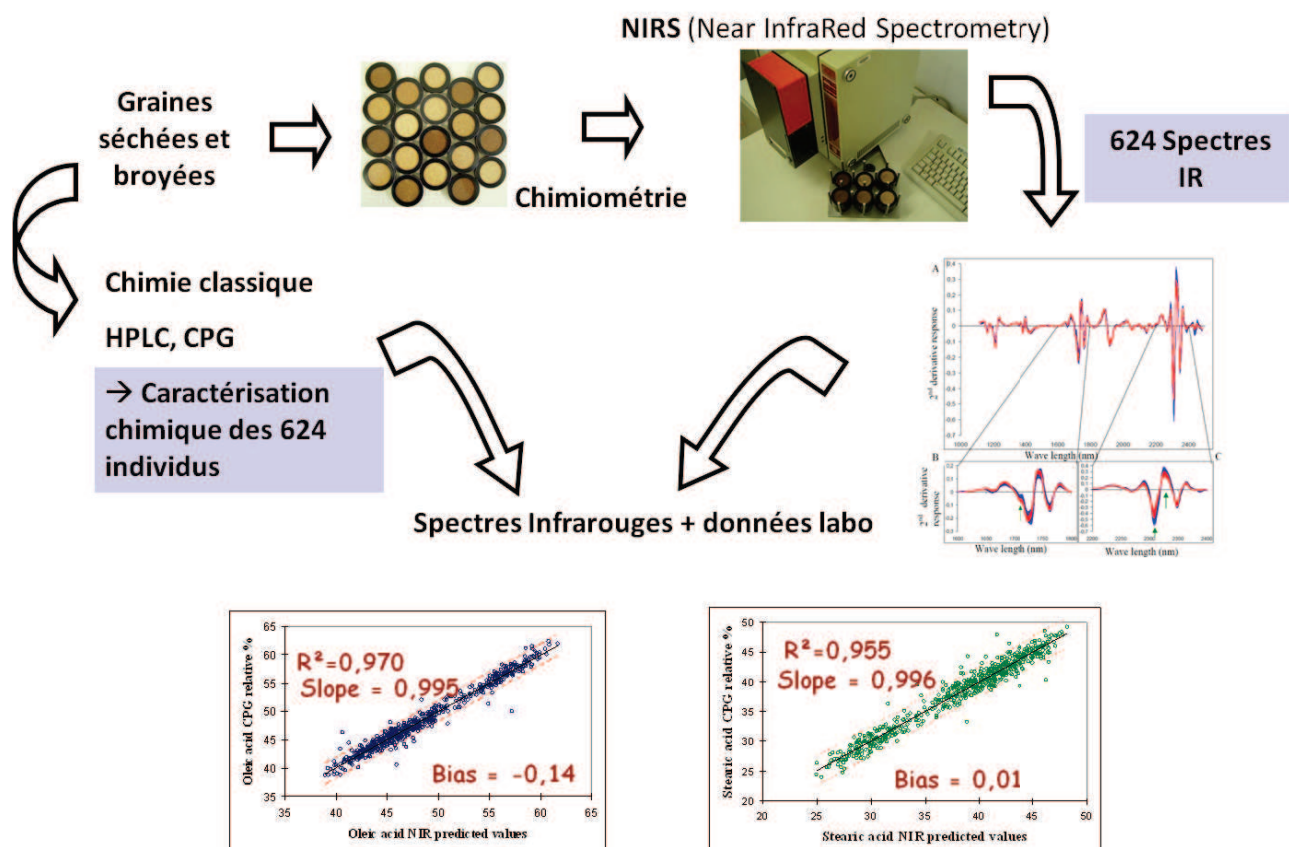
Pays	Site	Latitude	Longitude	Sous-espèce
Sénégal	Bandafassi	N 12.53	W 12.32	ssp. paradoxa
Sénégal	Samecoutha	N 12.60	W 12.13	ssp. paradoxa
Mali	Moroyafara	N 12.81	W 11.18	ssp. paradoxa
Mali	Mouralia	N 12.98	W 11.17	ssp. paradoxa
Mali	Mahinandi	N 13.78	W 10.85	ssp. paradoxa
Mali	Sonfara	N 13.24	W 10.47	ssp. paradoxa
Mali	Manako	N 13.11	W 9.45	ssp. paradoxa
Mali	Tenefina	N 12.92	W 8.97	ssp. paradoxa
Mali	Badougou	N 11.26	W 8.32	ssp. paradoxa
Mali	Filamana	N 10.51	W 7.97	ssp. paradoxa
Mali	Sotuba	N 12.66	W 7.92	ssp. paradoxa
Mali	Madina Diassa	N 10.82	W 7.74	ssp. paradoxa
Mali	Sido	N 11.67	W 7.60	ssp. paradoxa
Mali	Sirakorola	N 13.29	W 7.54	ssp. paradoxa
Mali	Tingole	N 12.75	W 7.16	ssp. paradoxa
Mali	Koumantou	N 11.40	W 6.85	ssp. paradoxa
Mali	BoubouWere	N 12.90	W 6.84	ssp. paradoxa
Mali	Kodougouni	N 12.97	W 6.66	ssp. paradoxa
Mali	massala	N 13.28	W 6.51	ssp. paradoxa
Mali	Daoulabougou	N 13.25	W 6.45	ssp. paradoxa
Mali	Banangoroni	N 13.32	W 6.39	ssp. paradoxa
Mali	Nafegue	N 10.51	W 5.98	ssp. paradoxa
Mali	Ouatiali	N 10.74	W 5.93	ssp. paradoxa
Mali	Zamblara	N 11.37	W 5.71	ssp. paradoxa
Mali	Pourou	N 10.58	W 5.71	ssp. paradoxa
Mali	Katiena	N 13.32	W 5.65	ssp. paradoxa
Mali	Danzana	N 12.57	W 5.63	ssp. paradoxa
Mali	M'Peresso	N 12.29	W 5.38	ssp. paradoxa
Mali	Daelan	N 13.25	W 4.99	ssp. paradoxa
Mali	Mansaba	N 13.92	W 4.25	ssp. paradoxa
Mali	Tori	N 13.60	W 3.72	ssp. paradoxa
Mali	Songo	N 14.40	W 3.70	ssp. paradoxa
Mali	Sassambourou	N 14.31	W 3.50	ssp. paradoxa
Mali	Tenndeli	N 14.09	W 3.45	ssp. paradoxa
Mali	Dourou	N 14.32	W 3.43	ssp. paradoxa
Mali	Koporo-Pen	N 14.24	W 3.27	ssp. paradoxa
Mali	Kiri	N 13.98	W 3.04	ssp. paradoxa
Mali	Bargou	N 14.02	W 2.98	ssp. paradoxa
Côte d'Ivoire	Napieoledougou	N 9.280	W 5.57	ssp. paradoxa
Côte d'Ivoire	Dolekaha	N 9.57	W 5.37	ssp. paradoxa
Côte d'Ivoire	Dassoumble	N 9.38	W 4.81	ssp. paradoxa
Burkina Faso	Niangoloko	N 10.31	W 4.93	ssp. paradoxa
Burkina Faso	Peni	N 10.97	W 4.45	ssp. paradoxa
Burkina Faso	Karo	N 12.36	W 3.30	ssp. paradoxa
Burkina Faso	Gobila	N 12.93	W 2.30	ssp. paradoxa
Burkina Faso	Titao	N 13.72	W 2.15	ssp. paradoxa
Burkina Faso	Guibare	N 13.06	W 1.60	ssp. paradoxa
Burkina Faso	Bonogo	N 12.06	W 1.60	ssp. paradoxa
Burkina Faso	Sapone	N 12.57	W 1.57	ssp. paradoxa
Burkina Faso	Pô	N 11.18	W 1.15	ssp. paradoxa

Pays	Site	Latitude	Longitude	Sous-espèce
Burkina Faso	Konia	N 13.10	W 0.99	ssp. paradoxa
Burkina Faso	Touroum	N 13.18	W 0.93	ssp. paradoxa
Burkina Faso	Fada N'Gourma	N 12.06	E 0.41	ssp. paradoxa
Ghana	Larabanga	N 9.19	W 1.85	ssp. paradoxa
Ghana	Mole	N 9.19	W 1.70	ssp. paradoxa
Ghana	Kawampe	N 8.44	W 1.57	ssp. paradoxa
Ghana	Zolanyili	N 9.43	W 1.00	ssp. paradoxa
Ghana	Sumbrungu	N 10.83	W 0.96	ssp. paradoxa
Ghana	Kadia	N 9.90	W 0.86	ssp. paradoxa
Ghana	Yendi	N 9.45	W 0.04	ssp. paradoxa
Togo	Tabango	N 10.74	E 0.19	ssp. paradoxa
Togo	Santioug	N 10.47	E 0.39	ssp. paradoxa
Togo	Mango	N 10.38	E 0.45	ssp. paradoxa
Togo	Payô	N 10.23	E 0.69	ssp. paradoxa
Togo	Koudassi	N 6.64	E 0.86	ssp. paradoxa
Togo	Kpasside	N 10.05	E 0.94	ssp. paradoxa
Togo	Kante	N 9.98	E 1.03	ssp. paradoxa
Togo	Bounoh	N 9.47	E 1.11	ssp. paradoxa
Togo	Yialoufon	N 8.96	E 1.12	ssp. paradoxa
Togo	Kolina	N 9.10	E 1.14	ssp. paradoxa
Togo	Djangbassou	N 7.81	E 1.19	ssp. paradoxa
Bénin	Dikouan	N 10.28	E 1.30	ssp. paradoxa
Bénin	Cocota	N 10.41	E 1.36	ssp. paradoxa
Bénin	Pendjari	N 11.05	E 1.37	ssp. paradoxa
Bénin	Kpale	N 9.65	E 1.73	ssp. paradoxa
Bénin	Pira	N 8.49	E 1.77	ssp. paradoxa
Bénin	Kandi	N 11.16	E 2.95	ssp. paradoxa
Bénin	Kalale	N 10.27	E 3.36	ssp. paradoxa
Niger	Boumba	N 12.41	E 2.86	ssp. paradoxa
Nigeria	Wawa	N 9.90	E 4.42	ssp. paradoxa
Nigeria	Rijau	N 11.10	E 5.25	ssp. paradoxa
Nigeria	Paiko	N 9.43	E 6.60	ssp. paradoxa
Nigeria	Abuja	N 9.08	E 7.42	ssp. paradoxa
Nigeria	Kuru	N 9.68	E 8.81	ssp. paradoxa
Nigeria	Shendam	N 8.88	E 9.53	ssp. paradoxa
Nigeria	Kaltungo	N 9.82	E 11.32	ssp. paradoxa
Nigeria	Mubi	N 10.27	E 13.27	ssp. paradoxa
Cameroun	Bangangte	N 5.15	E 10.52	ssp. paradoxa
Cameroun	Foumban	N 4.73	E 10.90	ssp. paradoxa
Cameroun	Ngong	N 9.02	E 13.50	ssp. paradoxa
Cameroun	Ngaoundere	N 7.32	E 13.58	ssp. paradoxa
Cameroun	Mafa-Kilda	N 9.03	E 13.75	ssp. paradoxa
Cameroun	Figuil	N 9.75	E 13.95	ssp. paradoxa
Cameroun	Mendeo	N 10.37	E 14.43	ssp. paradoxa
Tchad	Kelo	N 9.20	E 15.78	ssp. paradoxa
Tchad	Mondou	N 8.62	E 16.02	ssp. paradoxa
Tchad	Sarh	N 9.17	E 18.37	ssp. paradoxa
Rep. Centrafricaine	North	N 8.58	E 20.45	ssp. paradoxa
Central Afr. Rep.	Central	N 8.22	E 20.59	ssp. paradoxa
Central Afr. Rep.	South	N 7.88	E 20.85	ssp. paradoxa
Sudan	Kuajena	N 7.52	E 28.25	ssp. nilotica
Sudan	Mapel	N 7.29	E 28.45	ssp. nilotica
Sudan	Tonj	N 7.21	E 28.65	ssp. nilotica
Sudan	Cuebet	N 7.03	E 29.21	ssp. nilotica
Sudan	Wulu	N 6.49	E 29.63	ssp. nilotica
Sudan	Kuel Kuac	N 6.58	E 29.64	ssp. nilotica
Sudan	Matengai	N 6.83	E 29.64	ssp. nilotica
Sudan	Makembele	N 6.64	E 29.65	ssp. nilotica
Sudan	Malau	N 6.71	E 29.66	ssp. nilotica
Sudan	Abin Ajok	N 6.72	E 29.66	ssp. nilotica



Pays	Site	Latitude	Longitude	Sous-espèce
Sudan	Waco	N 6.18	E 29.67	ssp. nilotica
Sudan	Mvolo	N 6.03	E 29.92	ssp. nilotica
Sudan	Lado	N 6.01	E 30.03	ssp. nilotica
Sudan	Yeri	N 5.87	E 30.08	ssp. nilotica
Sudan	Mbara	N 5.45	E 30.17	ssp. nilotica
Sudan	Mundri	N 5.34	E 30.31	ssp. nilotica
Uganda	Arua	N 3.03	E 30.90	ssp. nilotica
Uganda	Eria	N 3.62	E 31.64	ssp. nilotica
Uganda	Adumara	N 2.53	E 33.05	ssp. nilotica
Uganda	Otongere	N 2.72	E 33.15	ssp. nilotica
Uganda	Okwongo	N 2.50	E 33.25	ssp. nilotica
Uganda	Lacankweri	N 2.80	E 33.31	ssp. nilotica
Uganda	Olwoko	N 2.43	E 33.50	ssp. nilotica
Uganda	Otuke ss	N 2.51	E 33.70	ssp. nilotica
Uganda	Abarilela	N 1.98	E 33.81	ssp. nilotica
Uganda	Akoboï	N 1.82	E 33.99	ssp. nilotica

ANNEXE III.1. DÉMARCHE POUR LE DÉVELOPPEMENT DES ÉQUATIONS DE PRÉDICTION PAR SPECTROSCOPIE PROCHE INFRAROUGE SPIR



ANNEXE III.2. ANOVAS ET COMPARAISON DE MOYENNES ENTRE RÉGIONS

SITES ET TRANSECT POUR TOUS LES CARACTÈRES CHIMIQUES ÉTUDIÉS

Sampling considered	Fat Content					Stearic acid					Oleic acid				
	ID	N	Mean (CV)	Group	F p	ID	N	Mean (CV)	Group	F p	ID	N	Mean (CV)	Group	F p
Natural Area	W	367	49.1 (10)	A	56.52 <0.01	E	89	29.7 (7)	A	790.8 <0.01	W	366	45.9 (7)	A	967.4 <0.01
	E	89	53.2 (5)	B		W	366	40.8 (9)	B		E	89	56.4 (3)	B	
Total Nat. A. ¹		455	49.9 (10)				455	38.6 (14)				455	47.9 (10)		
WEST	3	10	40.3 (17)	A	10.8 <0.01	5	37	36.8 (9)	A	9.1 <0.01	11	34	44.3 (5)	A	8.1 <0.01
	1	12	42.0 (9)	A		1	12	37.5 (8)	AB		12	35	44.3 (6)	A	
	5	37	47.4 (11)	B		3	10	39.7 (7)	ABC		2	18	44.3 (4)	AB	
	9	18	47.8 (9)	BC		8	35	40.2 (8)	BC		6	37	45.0 (9)	AB	
	2	18	47.9 (7)	BC		7	35	40.4 (7)	BC		13	34	45.1 (7)	AB	
	8	35	48.3 (12)	BC		13	34	40.9 (9)	BC		9	18	45.5 (6)	AB	
	11	34	48.8 (8)	BC		4	40	41.1 (7)	BC		10	21	45.5 (6)	AB	
	7	35	48.9 (5)	BC		9	18	41.4 (7)	BC		3	10	45.7 (4)	ABC	
	10	21	49.2 (7)	BCD		10	21	41.7 (7)	C		4	40	46.6 (6)	BC	
	13	35	49.4 (11)	BCD		6	37	41.8 (11)	C		8	35	46.8 (6)	BC	
	6	37	51.1 (8)	CD		2	18	42.3 (6)	C		7	35	46.9 (5)	BC	
	12	35	51.4 (8)	CD		12	35	42.7 (7)	C		1	12	47.7 (4)	BC	
	4	40	52.8 (5)	D		11	34	42.7 (6)	C		5	37	48.8 (6)	C	
Total West		367	49.1 (10)				366	40.8 (9)				366	45.9 (7)		
EAST	17	19	51.5 (4)	A	6.0 <0.01	16	16	28.6 (5)	A	7.2 <0.01	17	19	54.5 (4)	A	13.4 <0.01
	15	29	52.7 (6)	AB		14	25	29.1 (7)	A		15	29	56.3 (3)	B	
	14	25	54.1 (5)	B		15	29	29.8 (6)	AB		14	25	57.2 (3)	B	
	16	16	54.8 (2)	B		17	19	31.1 (6)	B		16	16	57.5 (2)	B	
Total East		89	53.2 (5)				89	29.7 (7)				89	56.4 (3)		
Mali Transect	5	37	47.4 (11)	A	9.9 <0.01	5	37	36.8 (9)	A	12.1 <0.01	6	37	45.0 (9)	A	8.1 <0.01
	8	35	48.3 (12)	AB		8	35	40.2 (8)	B		4	40	46.6 (6)	A	
	7	35	48.9 (5)	AB		7	35	40.4 (7)	B		8	35	46.8 (6)	A	
	6	37	51.1 (8)	BC		4	40	41.1 (7)	B		7	35	46.9 (5)	A	
	4	40	52.8 (5)	C		6	37	41.8 (11)	B		5	37	48.8 (6)	B	
Total M. Trans. ²		184	49.8 (9)				187	40.1 (9)				187	46.8 (7)		
Ghana-Burkina Transect	9	18	47.8 (9)	A	2.7 <0.01	13	34	40.9 (9)	A	2.3 0.06	11	34	44.3 (5)	A	1.4 0.22
	11	34	48.8 (8)	AB		9	18	41.4 (7)	A		12	35	44.3 (6)	A	
	10	21	49.2 (7)	AB		10	21	41.7 (7)	A		13	34	45.1 (7)	A	
	13	35	49.4 (11)	AB		12	35	42.7 (7)	A		9	18	45.5 (6)	A	
	12	35	51.4 (8)	B		11	34	42.7 (6)	A		10	21	45.5 (6)	A	
Total G.-B. Trans. ³		143	49.5 (9)				142	42.0 (7)				142	44.8 (6)		

¹: Total Natural Area; ²: Total Mali transect; ³: Total Ghana-Burkina transect; N: number of trees analyzed; Mean: average value; CV: coefficient of variation; Group: grouping proposed by Bonferroni's test; p: p-value; W: West region; E: East region.

Annexe III.2. continues

Sampling considered	Palmitic acid					Vaccenic acid					Linoleic acid				
	ID	N	Mean (CV)	Group	F p	ID	N	Mean (CV)	Group	F p	ID	N	Mean (CV)	Group	F p
Natural Area	W	366	4.3 (18)	A	35.8	W	366	0.4 (43)	A	83.2	W	366	7.1 (15)	A	7.9
	E	89	4.8 (9)	B	<0.01	E	89	0.5 (32)	B	<0.01	E	89	7.3 (10)	A	0.17
<i>Total Nat. A.¹</i>		455	4.4 (17)				455	0.4 (44)				455	7.1 (14)		
WEST	4	40	3.6 (12)	A	23.3 <0.01	7	35	0.3 (38)	A	7.5 <0.01	2	18	6.4 (18)	A	4.5 <0.01
	7	35	3.9 (15)	AB		4	40	0.3 (45)	A		11	34	6.7 (15)	AB	
	10	21	3.9 (9)	AB		11	34	0.3 (56)	A		3	10	6.8 (18)	ABC	
	9	18	4.0 (11)	ABC		6	37	0.3 (47)	A		10	21	6.8 (11)	ABC	
	8	35	4.1 (16)	ABC		2	18	0.3 (43)	AB		12	35	6.9 (10)	ABC	
	11	34	4.2 (11)	BC		10	21	0.4 (29)	AB		6	37	7.0 (15)	ABC	
	12	35	4.3 (13)	BC		8	35	0.4 (33)	AB		4	40	7.0 (10)	ABC	
	6	37	4.3 (13)	BC		9	18	0.4 (55)	AB		9	18	7.0 (11)	ABC	
	5	37	4.3 (15)	BC		5	37	0.4 (35)	B		7	35	7.1 (13)	ABC	
	13	34	4.5 (13)	C		13	34	0.4 (34)	B		1	12	7.1 (17)	ABC	
	2	18	5.3 (17)	D		12	35	0.5 (31)	B		8	35	7.1 (20)	ABC	
	3	10	5.8 (18)	D		3	10	0.5 (41)	B		13	34	7.5 (15)	BC	
	1	12	5.8 (6)	D		1	12	0.5 (19)	B		5	37	8.0 (15)	C	
<i>Total West</i>		366	4.3 (18)				366	0.4 (43)				366	7.1 (15)		
EAST	15	29	4.6 (7)	A	6.5 <0.01	14	25	0.5 (48)	A	3.9 0.01	16	16	7.0 (10)	A	0.9 0.43
	14	25	4.7 (9)	A		15	29	0.5 (26)	AB		14	25	7.3 (10)	A	
	16	16	4.9 (8)	AB		16	16	0.6 (19)	AB		17	19	7.3 (6)	A	
	17	19	5.1 (9)	B		17	19	0.6 (24)	B		15	29	7.4 (11)	A	
<i>Total East</i>		89	4.8 (9)				89	0.5 (32)				89	7.2 (9)		
Mali Transect	4	40	3.6 (12)	A	10.2 <0.01	7	35	0.3 (38)	A	10.1 <0.01	6	37	7.0 (15)	A	6.2 <0.01
	7	35	3.9 (15)	AB		4	40	0.3 (45)	AB		4	40	7.0 (10)	A	
	8	35	4.1 (16)	BC		6	37	0.3 (47)	AB		7	35	7.1 (13)	A	
	6	37	4.3 (13)	C		8	35	0.4 (33)	BC		8	35	7.1 (20)	A	
	5	37	4.3 (15)	C		5	37	0.4 (35)	C		5	37	8.0 (15)	B	
<i>Total M. Trans.²</i>		184	4.1 (16)				187	0.3 (43)				187	7.2 (16)		
Ghana- Burkina Transect	10	21	3.9 (9)	A	5.1 <0.01	11	34	0.3 (56)	A	5.8 <0.01	11	34	6.7 (15)	A	3.8 <0.01
	9	18	4.0 (11)	A		10	21	0.4 (29)	AB		10	21	6.8 (11)	A	
	11	34	4.2 (11)	AB		9	18	0.4 (55)	AB		12	35	6.9 (10)	AB	
	12	35	4.3 (13)	AB		13	34	0.4 (34)	B		9	18	7.0 (11)	AB	
	13	34	4.5 (13)	B		12	35	0.5 (31)	B		13	34	7.5 (15)	B	
<i>Total G.-B. Trans.³</i>		142	4.2 (12)				142	0.4 (42)				142	7.0 (14)		

¹: Total Natural Area; ²: Total Mali transect; ³: Total Ghana-Burkina transect; N: number of trees analyzed; Mean: average value; CV: coefficient of variation; Group: grouping proposed by Bonferroni's test; p: p-value; W: West region; E: East region.

Annexe III.2. continues

Sampling considered	Linolenic acid					Arachidic acid					SU ratio				
	ID	N	Mean (CV)	Group	F <i>p</i>	ID	N	Mean (CV)	Group	F <i>p</i>	ID	N	Mean (CV)	Group	F <i>p</i>
Natural Area	W	366	0.3 (60)	B	73.6	E	89	1.0 (11)	A	267.8	E	89	0.55 (9)	A	691.1
	E	89	0.4 (17)	A	<0.01	W	366	1.2 (13)	B	<0.01	W	366	0.87 (14)	B	<0.01
<i>Total Nat. A.¹</i>		<i>455</i>	<i>0.3 (54)</i>				<i>455</i>	<i>1.2 (16)</i>				<i>455</i>	<i>0.81 (21)</i>		
WEST	8	35	0.1 (118)	A	18.3 <0.01	1	12	1.1 (6)	A	11.9 <0.01	5	37	0.74 (13)	A	8.7 <0.01
	4	40	0.2 (84)	AB		12	35	1.1 (10)	A		1	12	0.80 (11)	AB	
	7	35	0.2 (66)	ABC		13	34	1.1 (11)	A		7	35	0.84 (10)	BC	
	6	37	0.2 (68)	ABC		3	10	1.2 (10)	AB		8	35	0.84 (14)	BC	
	5	37	0.2 (61)	ABCD		5	37	1.2 (9)	AB		4	40	0.85 (11)	BCD	
	2	18	0.3 (50)	BCDE		9	18	1.2 (12)	AB		3	10	0.88 (11)	BCD	
	3	10	0.3 (24)	CDE		2	18	1.2 (12)	AB		9	18	0.88 (12)	BCD	
	1	12	0.4 (37)	DE		4	40	1.2 (11)	AB		13	34	0.88 (14)	BCD	
	11	34	0.4 (36)	E		8	35	1.3 (13)	B		10	21	0.89 (11)	BCD	
	9	18	0.4 (49)	E		11	34	1.3 (11)	B		6	37	0.92 (17)	BCD	
	13	34	0.4 (34)	E		10	21	1.3 (10)	BC		12	35	0.93 (12)	CD	
	12	35	0.4 (19)	E		7	35	1.3 (11)	BC		11	34	0.94 (10)	D	
	10	21	0.4 (32)	E		6	37	1.4 (11)	C		2	18	0.96 (11)	D	
<i>Total West</i>		<i>366</i>	<i>0.3 (60)</i>				<i>366</i>	<i>1.2 (13)</i>				<i>366</i>	<i>0.87 (14)</i>		
EAST	16	16	0.4 (12)	A	0.72 0.54	16	16	0.9 (14)	A	1.3 0.27	14	25	0.53 (9)	A	8.8 <0.01
	17	19	0.4 (11)	A		14	25	1.0 (11)	A		16	16	0.53 (7)	A	
	15	29	0.4 (24)	A		15	29	1.0 (11)	A		15	29	0.55 (8)	A	
	14	25	0.4 (12)	A		17	19	1.0 (9)	A		17	19	0.59 (8)	B	
<i>Total East</i>		<i>89</i>	<i>0.4 (17)</i>				<i>89</i>	<i>1.0 (11)</i>				<i>89</i>	<i>0.55 (9)</i>		
Mali Transect	8	35	0.1 (118)	A	2.2 0.07	5	37	1.2 (9)	A	10.8 <0.01	5	37	0.74 (13)	A	11.6 <0.01
	4	40	0.2 (84)	A		4	40	1.2 (11)	AB		7	35	0.84 (10)	B	
	7	35	0.2 (66)	A		8	35	1.3 (13)	AB		8	35	0.84 (14)	B	
	6	37	0.2 (68)	A		7	35	1.3 (11)	BC		4	40	0.85 (11)	B	
	5	37	0.2 (61)	A		6	37	1.4 (11)	C		6	37	0.92 (17)	B	
<i>Total M. Trans.²</i>		<i>184</i>	<i>0.2 (77)</i>				<i>187</i>	<i>1.3 (12)</i>				<i>187</i>	<i>0.84 (15)</i>		
Ghana-Burkina Transect	11	34	0.4 (36)	A	0.2 0.95	12	35	1.1 (10)	A	12.5 <0.01	9	18	0.88 (12)	A	2.2 0.08
	9	18	0.4 (49)	A		13	34	1.1 (11)	A		13	34	0.88 (14)	A	
	13	34	0.4 (34)	A		9	18	1.2 (12)	AB		10	21	0.89 (11)	A	
	12	35	0.4 (19)	A		11	34	1.3 (11)	B		12	35	0.93 (12)	A	
	10	21	0.4 (32)	A		10	21	1.3 (10)	B		11	34	0.94 (10)	A	
<i>Total G.-B. Trans.³</i>		<i>142</i>	<i>0.4 (33)</i>				<i>142</i>	<i>1.2 (12)</i>				<i>142</i>	<i>0.91 (12)</i>		

¹: Total Natural Area; ²: Total Mali transect; ³: Total Ghana-Burkina transect; N: number of trees analyzed; Mean: average value; CV: coefficient of variation; Group: grouping proposed by Bonferroni's test; *p*: *p*-value; W: West region; E: East region.

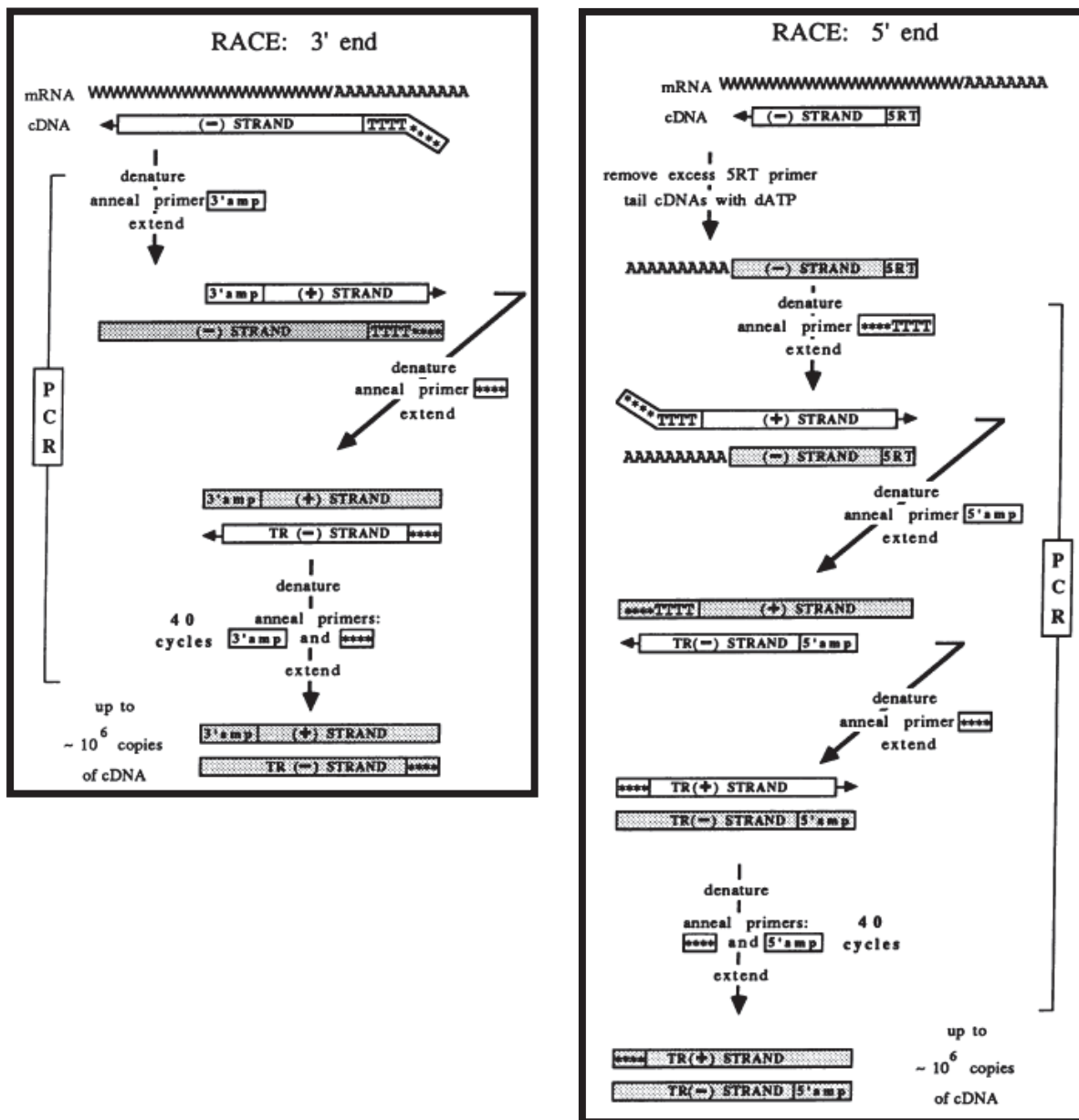
Annexe III.2. continues

Sampling considered	α -tocopherol					γ -tocopherol				
	ID	N	Mean (CV)	Group	F p	ID	N	Mean (CV)	Group	F p
Natural Area	W	246	110.6 (29)	A	3.1	W	246	10.5 (61)	B	267.6
	E	27	122.0 (22)	A	0.08	E	27	33.7 (32)	A	<0.01
<i>Total Nat. A.</i> ¹		273	111.7 (29)				273	12.8 (77)		
WEST	1	4	54.9 (53)	A	9.7 <0.01	6	37	6.9 (56)	A	2.8 0.02
	2	9	74.7 (44)	A		11	9	7.6 (43)	A	
	6	37	94.3 (27)	AB		7	35	8.1 (45)	A	
	4	40	96.6 (20)	AB		10	7	9.4 (39)	A	
	13	10	98.8 (24)	ABC		4	40	11.7 (75)	A	
	11	9	101.7 (21)	ABC		12	9	11.8 (74)	A	
	12	9	110.9 (14)	ABCD		5	37	11.8 (49)	A	
	7	35	112.7 (23)	BCD		8	35	12.0 (53)	A	
	8	35	121.4 (25)	CD		13	10	12.2 (54)	A	
	3	7	128.0 (44)	CD		2	9	12.6 (81)	A	
	10	7	130.0 (15)	CD		3	7	12.9 (36)	A	
	9	7	134.4 (17)	CD		9	7	14.4 (36)	A	
	5	37	138.5 (21)	D		1	4	14.5 (30)	A	
<i>Total West</i>		246	110.6 (29)				246	10.5 (61)		
EAST	16	4	104.1 (19)	A	2.8 0.06	15	9	24.1 (19)	A	7.0 <0.01
	15	9	111.0 (30)	A		14	9	35.3 (34)	AB	
	14	9	129.4 (14)	A		17	5	41.5 (14)	B	
	17	5	143.0 (11)	A		16	4	42.0 (13)	B	
<i>Total East</i>		27	122.0 (22)				27	33.7 (31)		
Mali Transect	6	37	94.3 (27)	A	18.1 <0.01	6	37	6.9 (56)	A	5.9 <0.01
	4	40	96.6 (20)	AB		7	35	8.1 (45)	AB	
	7	35	112.7 (23)	BC		4	40	11.7 (75)	B	
	8	35	121.4 (25)	CD		5	37	11.8 (49)	B	
	5	37	138.5 (21)	D		8	35	12.0 (53)	B	
<i>Total M. trans.</i> ²		184	112.3 (27)				184	10.1 (63)		
Ghana-Burkina Transect	13	10	98.8 (24)	A	4.9 <0.01	13	10	7.6 (54)	A	1.6 0.21
	11	9	101.7 (21)	AB		10	7	9.4 (39)	A	
	12	9	110.9 (14)	ABC		12	9	11.8 (74)	A	
	10	7	130.0 (15)	BC		11	9	12.2 (43)	A	
	9	7	134.4 (17)	C		9	7	14.4 (36)	A	
<i>Total G.-B. trans.</i> ³		42	113.1 (21)				42	11.0 (55)		

¹: Total Natural Area; ²: Total Mali transect; ³: Total Ghana-Burkina transect; N: number of trees analyzed; Mean: average value; CV: coefficient of variation; Group: grouping proposed by Bonferroni's test; p: p-value; W: West region; E: East region.

ANNEXE IV.1. RACE (RAPID AMPLIFICATION OF cDNA ENDS)

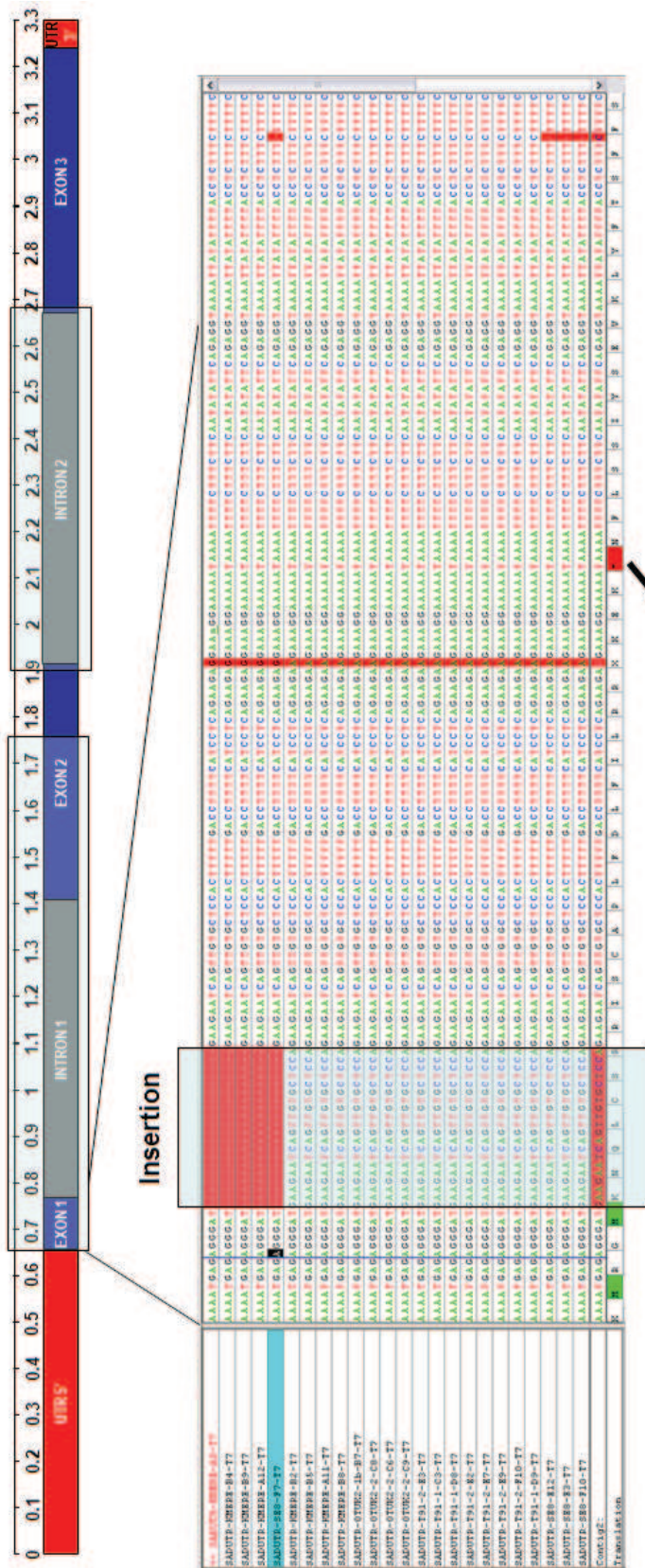
Représentation schématique tirée de la publication originale de Frohman *et al.* (1988)



Primers: ****TTTT (dT)17-adaptor, 5'-GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTT-3'. This sequence contains the Xho I, Sal I, and Cla I recognition sites. ****, Adaptor, 5'-GACTCGAGTCGACATCG-3'. 3'amp (amp refers to amplification), specific to gene of interest, complementary to (-) strand. 5RT (RT refers to reverse transcription) and 5'amp, specific to gene of interest, complementary to (+) strand. Open rectangles represent DNA strands actively being synthesized; shaded rectangles represent DNA previously synthesized. At each step the diagram is simplified to illustrate only how the new product formed during the previous step is utilized. A (-) or (+) strand is designated as "truncated" (TR) when it is shorter than the original (-) or (+) strand, respectively.

ANNEXE IV.2 : ALIGNEMENT DE QUELQUES SÉQUENCES *SAD2* POUR LES

SOU



In silico translation shows a truncated protein for 1 of the copy : Stop codon in exon 1

But not in subspecies *nilotica* in which *SAD2* is still functional

